# TRANSLATION IN WIKIPEDIA: A PRAXEOLOGICAL STUDY OF NORMATIVITY, NEGOTIATION AND AUTOMATION ACROSS FOUR LANGUAGE COMMUNITIES

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Humanities

**2021**

**JOSÉ G. GÓNGORA-GOLOUBINTSEFF**
**SCHOOL OF ARTS, LANGUAGES AND CULTURES**

# Table of contents

Word count: 77,357

# List of figures and tables

# Abstract

Launched in 2001, Wikipedia is a long-standing multilingual user-driven encyclopaedia and one of the most popular sources of first-hand information in the world. Although its previously neglected multilingual nature has awakened scholarly interest in recent years, most studies have overlooked the standards and materials that underpin and configure the practice of translation in Wikipedia. Consequently, this doctoral thesis first sets out to investigate the extent to which translation standards have been regulated, negotiated and ultimately incorporated into the practices of 16 translators of the Spanish, French, Dutch and Swedish language communities of the encyclopaedia. This investigation is then followed by an examination of the translators' views on and deployment of automated devices such as software robots (bots) and Wikipedia's bespoke Content Translation Tool (CX). Drawing primarily upon Wenger's (1998) 'communities of practice', Warde's (2016) 'standards of performance', and Shove's (2017) concept of 'devices as configuring elements' of practice, this project seeks to ascertain the role of local (community-based) translation standards and automated devices in configuring the practice of translation in Wikipedia, with the focus placed on the last lustrum.

The thematic analysis of documented standards revealed that despite there being tangible differences in how the four communities regulated translation, most guidelines gave similar advice on core editing principles such as verifiability of the sources. The data collected from translation-related comments on the ancillary 'talk pages' further suggest that whilst certain aspects of the standards have been vehemently contested, such documents have not undergone substantial changes over the years. This stagnation was later corroborated by a cohort of experienced translators who took part in the study, most of whom attached little, if any, relevance to local standards. The findings also show a widespread tendency among participants to comply with more 'enforceable' policies commonly found in editing, thus lending support to previously formulated claims that translation and editing in Wikipedia form a continuum. Finally, the study has also brought to light the impact of devices such as bots and CX in reconfiguring translation in the encyclopaedia. Specifically, it has shown that, although the two devices were created to facilitate the dissemination of knowledge across language communities, an overreliance on bots has contributed to increasing the workload of editor-translators, whereas CX has optimised their productivity.

# Declaration and Copyright statement

I declare that that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

**i.** The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the 'Copyright') and s/he has given the University of Manchester certain rights to use such Copyright, including for administrative purposes.

**ii.** Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

**iii.** The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the 'Intellectual Property') and any reproductions of copyright works in the thesis, for example graphs and tables ('Reproductions'), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

**iv.** Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, the University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in the University's policy on Presentation of Theses.

# Acknowledgements

First of all, I would like to thank my supervisor, Professor Maeve Olohan, for her invaluable feedback, meticulous and thought-provoking observations, pastoral support, and practical advice throughout the past three years. This project would not have come to fruition without her guidance and encouragement. I would also like to extend my gratitude to the other two members of the supervisory team, Dr Julio Villa-García and Professor Luis Pérez-González, whose insightful comments helped me reconsider aspects of my study.

Second, I would like to voice my appreciation to Dr Jan Buts for his timely assistance with the Dutch-to-English translations of the Wikipedia guidelines. In a similar vein, I want to acknowledge the altruistic collaboration of the Wikipedia translators that took part in the study. Their stories were indeed inspiring and deserve more recognition.

Third, I would like to give thanks to my friend and colleague, Néstor Singer, whose joyful personality, hard work and unwavering ability to grow through adversity have been a continuous source of inspiration to me during our shared doctoral journey. I am confident our professional and personal relationship will continue to flourish well into the future. I am also indebted to fellow researchers Chris Rhodes and Chit Fung (Lawrence) Lam, whose friendship and kindness contributed significantly to improving my wellbeing at the height of the pandemic.

Lastly, I am grateful to my dearest partner, whom I met in Manchester, my parents, siblings, and grandmothers. Despite the geographical barriers and current restrictions on international travel, various family members have managed to guide me wisely and encourage me from afar.

This research project (Ref: 2068270) was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) as part of the Doctoral Training Partnership scheme.

# Chapter 1. Introduction

Launched in 2001 by Jimmy Wales and Larry Sanger, Wikipedia is probably best known for being a multilingual user-driven encyclopaedia 'that anyone can edit'. At the time of writing, Wikipedia is available in 323 languages and language varieties ('Wikipedia' 2021). Of these, the English-language version is by far the largest with ca. 6.5 million articles, followed by the mostly bot-generated Cebuano (5.5M) and Swedish (3.5M) editions. As a collaborative platform, Wikipedia relies upon the goodwill of thousands of volunteers, variously referred to as 'users', 'editors' or 'Wikipedians'.

One of Wikipedia's most salient features is that volunteers are anonymous. As (O'Sullivan 2009, 88) observes, anonymity in Wikipedia is of paramount importance because 'it allows all contributions to be judged on their intrinsic merit rather than by their source'. Therefore, to participate in the encyclopaedia, it is not necessary to create an account, and even those volunteers who opt to do so are not required to share their personal data (O'Sullivan 2009; Jones 2017). In practice, this gives Wikipedia editors leeway to create their own identity in the project without fear of being discriminated against because of their background or interests.

Despite this scenario, Wikipedia is far from egalitarian. The user-generated encyclopaedia is subject to the same flaws that are commonly observed in other online platforms. Various studies have tackled lingering problems such as cultural bias (Callahan and Herring 2011; Townsend, Osmond, and Phillips 2013), the seeming lack of neutrality across Wikipedia language communities (Young, Wigdor, and Kane 2020), and the gender gap (Shane-Simpson and Gillespie-Lynch 2017). For instance, a recent study found that female administrators or system operators ('sysops') in Wikipedia tend to converge to the behaviour of their male colleagues (Gallus and Bhatia 2020).

Besides issues of gender and neutrality, Wikipedia has sometimes been perceived as a hierarchical environment, where veteran editors can often be strict and dogmatic (Mattus 2014; Khazraie and Talebzadeh 2020). According to Mattus (2014), such editors cling to their powerful positions and 'may prevent other voices from being heard'. This hostility towards disempowered new members has not gone

unnoticed by Wikipedia's co-founder Jimmy Wales. The online encyclopaedia not only has behavioural guidelines to tackle harassment (‘Wikipedia:Harassment’ 2021), but Wales made it clear in his ‘statement of principles’ (Wales 2021) that:

> Newcomers are always to be welcomed. There must be no cabal, there must be no elites, there must be no hierarchy or structure which gets in the way of this openness to newcomers.

Senior editors, especially administrators, are often tasked with overseeing compliance with Wikipedia's standards and other aspects of governance. Although the encyclopaedia has a host of strict policies and guidelines for editors to enforce upon non-compliant users, previous research has shown that regulation is often achieved without invoking the rules (Goldspink 2010). Regarding the communal aspects of norm enforcement, some authors found that competent editors, regardless of their seniority status, are less likely to be on the ‘watchlist’ than beginners (Klapper and Reitzig 2018). Similarly, seniority status within a given Wikipedia community has been noted to be irrelevant to non-members (Lanamäki and Lindman 2018).

The reasons that lead individuals to join Wikipedia have also received scholarly attention. O'Sullivan (2009) observes that a sense of belonging and contributing to something of significant proportions such as Wikipedia may encourage some people to volunteer. The encyclopaedia has a series of intangible, non-monetary rewards that editors can allocate to one another in recognition of substantial contributions to knowledge. As subjective as it may seem, this positive reinforcement validates the editors’ efforts and brings long-term benefits to the encyclopaedia. Townsend, Osmond, and Phillips (2013) note that ‘this reward system represents the gamification of knowledge in Wikipedia’. Likewise, research by Pee (2018) has demonstrated that individuals who contribute to Wikipedia do so through a relationship of interdependence, i.e. taking into account others’ needs.

Although research on the collaborative processes of knowledge building in Wikipedia has traditionally had the English version as the principal object of enquiry, the last decade has seen a surge of academic interest in the multilingual potential of the encyclopaedia. Such interest was partly motivated by increasing awareness of the cross-cultural differences within and across individual language communities. As Hara, Shachaf, and Hew (2010) observe, shifting the attention to

other language versions of Wikipedia brings in a comparative perspective that contributes to a better understanding of how distinct communities are organised. In a similar vein, Callahan and Herring (2011) note that, while the English Wikipedia is likely to remain a global repository of human knowledge, analysing other language communities may help unpack their local values.

Against this backdrop, the following section begins by discussing research on Wikipedia's multilingualism, with the emphasis placed on cross-lingual collaboration and content creation. In line with this multilingual framework, subsection 1.1.1 goes on to provide an overview of prior studies on Wikipedia translation. In doing so, it will also address the difficulties of disentangling translation from other forms of content creation in Wikipedia. Finally, 1.1.2 presents research on attitudes towards automation in the user-generated encyclopaedia. When approaching automation, attention will be paid to Wikipedia's Content Translation Tool (CX) and the use of software robots (bots) as devices that enable the performance of a variety of tasks, including cross-lingual content creation.

## 1.1 Wikipedia's multilingualism

According to Wikipedia's co-founder Jimmy Wales, the 'prime objective' of the platform is 'to create and distribute a free encyclopaedia of the highest possible quality to every single person on the planet in their own language' ('Wikipedia:Prime Objective' 2021). Therefore, it is revealing that so little research has been undertaken on Wikipedia's multilingualism since its inception in 2001 (de Melo and Weikum 2014; Jones 2018b). In the 2008 book *How Wikipedia Works*, Phoebe Ayers, a well-known librarian and long-time Wikipedian, and colleagues dedicate merely one chapter out of seventeen to discuss the over 200 language communities at the time. They acknowledge the benefits of joining any of those communities by highlighting that, unlike the English version, smaller Wikipedias have a shortage of encyclopaedic content and thus require more volunteers (Ayers, Matthews, and Yates 2008).

Regardless of its briefness, their chapter on multilingualism provides an early insight into how individual communities may have their own set of established rules.

Moreover, they note that such regulations – in the form of policies and guidelines – 'are not dictated from on high' but, instead, 'they have been developed collaboratively by community members' (Ayers, Matthews, and Yates 2008, 363). On this premise, they insist that they can be negotiated, changed, and sometimes even ignored. They are cautious on this last point and opine that even though 'Wikipedia is not strictly governed by written rules', 'policies and guidelines [in the encyclopaedia] exist to create the best site possible' (Ayers, Matthews, and Yates 2008, 365). As such, they believe that they should only be circumvented if they interfere with common sense.

Much research on Wikipedia's multilingualism has concentrated on cultural diversity across communities. This diversity has led some scholars to question the suitability of using the term 'community' loosely when approaching Wikipedia. Ensslin (2011), for instance, argues that the encyclopaedia consists of 'a highly diversified range of users' bound together through a process of 'pseudo-unifying collectivisation' (Ensslin 2011, 545). She posits that 'community' is a reified or imaginary construct intended to imbue a sense of belonging among Wikipedia volunteers. In her view, individual language communities of the user-generated encyclopaedia are 'discursive spaces' characterised by 'local rules and restrictions' (Ensslin 2011, 554). Sections 1.2 and 2.2 will return to this idea when framing the study of Wikipedia as a community of practice.

Most studies on multilingual diversity in Wikipedia have focused on two main aspects: encyclopaedic content and cultural differences across communities. Regarding content, research in this area has shown that the interpretation of Wikipedia policies such as the 'Neutral point of view' (NPOV) and 'Notability' varies considerably across individual language communities (Callahan and Herring 2011; Góngora-Goloubintseff 2020; Park et al. 2020). For example, Callahan and Herring (2011) found notable differences in how famous Americans and Poles were portrayed in the English and Polish Wikipedia communities. Góngora-Goloubintseff (2020) identified cross-lingual neutrality breaches in a comparative case study of the English and Spanish Wikipedia articles on the Falklands War. Similarly, Park et al. (2020) unearthed systemic bias in LGBT narratives across the English, Russian and Spanish Wikipedia communities.

Along with the issues discussed above, other areas of concern have been the quality and reliability of the content. According to van Dijk (2009, 234), some Wikipedia language communities 'embellish the total number of articles by creating pseudo-articles with little or no encyclopaedic value'. Likewise, Lewoniewski, Węcel, and Abramowicz (2018) and Lewoniewski (2019) found evidence that standards of excellence in Wikipedia articles or entries depend on the topic and the language community involved.

As mentioned at the beginning of this section, debate has prevailed as to whether there are notable cultural differences across Wikipedia language communities. One example is Pfeil, Zaphiris, and Ang's (2006) examination of the relationship between national culture and editing patterns in the online encyclopaedia. They drew data from the Dutch, French, German and Japanese Wikipedia versions of the entry for 'game' and analysed the volunteers' contributions against Hofstede's cultural influence framework, which continues to be applied in many disciplinary contexts despite its rather reductionist treatment of culture.[1] Pfeil, Zaphiris, and Ang (2006) found significant differences across the four communities. For instance, they observed that Japanese Wikipedia editors were more reliant on policies and guidelines than editors of the other communities. Pfeil, Zaphiris, and Ang (2006) linked this stricter observance of the rules to Japan's higher Uncertainty Avoidance index and concluded that the cultural diversity ascertained in Wikipedia mirrored the physical world. By the same token, Nemoto and Gloor (2011) investigated culturally-influenced editing patterns across Wikipedia language communities. They found that communication among Japanese Wikipedia editors had hierarchical overtones, whereas interaction among their Finnish Wikipedia counterparts was substantially more egalitarian.

Moving from user interaction to editing, Samoilenko et al. (2016, 17) conducted quantitative research on diverse Wikipedia languages and came to the conclusion that each community of the encyclopaedia 'present[ed] a cultural memory place' with a unique repository of articles. Chandra and Maiti (2018) identified distinct editing patterns across Wikipedia communities. They discovered that senior editors in the English, French and Spanish language communities of Wikipedia tended to

[1] The four dimensions proposed by Hofstede (1991) are Power Distance, Collectivism versus Individualism, Femininity versus Masculinity, and Uncertainty Avoidance.

focus on articles of their interest. In contrast, their Swedish, Italian, Polish and Russian Wikipedia peers showed no preference.

Thus far, studies on Wikipedia's multilingual nature have tackled editing and behavioural patterns across communities. Although questions have been raised about cultural differences impacting content creation, studies on translation are relatively scarce for an encyclopaedia as linguistically diverse and long-lived as Wikipedia. Furthermore, translation scholars have yet to explore the role of automated devices in Wikipedia. As the following subsections will show, barring a few notable exceptions (McDonough Dolmaya 2017), content-creation devices such as bots and CX have received little attention in translation studies. Consequently, this thesis aims to investigate the use of automation as an essential component in aiding – and potentially configuring – the practice of translation in Wikipedia, alongside standards.

### 1.1.1 Wikipedia and translation studies

The phenomenon of volunteer translation (Pérez-González and Susam-Saraeva 2012) or commons-based peer production (Jiménez-Crespo 2017), as opposed to professional or remunerated translation, predates the Internet. Olohan (2012) posits that as early as the 19th century, the British periodical *Scientific Memoirs* survived for 15 years thanks to generous donations and the selfless effort of a community of volunteer translators possibly driven by a desire 'to expand scientific knowledge' and enhance their social standing. Such motivations could equally apply to the philosophy permeating Wikipedia. This public-spirited project has been in a constant quest for a world where everyone has 'free access to the sum of all human knowledge'('Wikipedia:Prime Objective' 2021).

Despite the fact that volunteer translation is not a new phenomenon, research on Wikipedia has been modest and relatively recent (Shuttleworth 2017; Jones 2018b). Early examples include Hautasaari et al.'s (2011) analysis of community behaviour in collaborative cross-wiki translation and McDonough Dolmaya's (2012) investigation of Wikipedia translators' profiles. The work of the latter, in particular, would influence future scholarship, as it pioneered the study of the ethos and motivations of Wikipedia translators. McDonough Dolmaya (2012) examined the

encyclopaedia translators' background and focused on their reasons to engage in what she refers to as an altruistic or cause-driven endeavour. To this end, she circulated an online survey among 204 English Wikipedia editors with substantial translation experience. Of 75 respondents, 89.3% were primarily interested in making information available to other language speakers (McDonough Dolmaya 2012, 182). Over half of the respondents (56%) reported that they were driven by a desire to support Wikipedia's aims, while 36% expressed an interest in practising their source language. Of note is that 68% of the participants declared that they did not have previous translation experience, and only a reduced number (15%) were involved in translation-related activities outside Wikipedia. Although McDonough Dolmaya (2012) only surveyed English Wikipedia translators, her study laid the groundwork for future research in the discipline (Olohan 2014; Shuttleworth 2017; Jones 2018a; Torres-Simón 2019).

Olohan (2014), for example, investigated the motivations of volunteer translators working for the American media organisation TED. In her analysis of 11 translators' blog posts, she found that the vast majority were driven by an altruistic desire to support TED's mission of supplying global knowledge, sharing ideas and effecting social change (Olohan 2014, 25). Similarly, Cámara de la Fuente (2015) provided important insight into the motivations of TED translators. Drawing on data collected from 177 self-administered structured questionnaires, she established that TED translators were primarily keen on contributing to TED's mission. The findings obtained in both studies accord with earlier observations, particularly with those of O'Brien and Schäler (2010), whose research centred on the motivations of volunteer translators in the not-for-profit Rosetta Foundation. After examining 139 answers to a questionnaire that used a Likert-scale rating, they discovered that most translators had an interest in supporting the Rosetta Foundation's cause, followed by a need to gain professional translation experience.

Returning to Wikipedia, other studies have explored its potential as a tool for translators (Alonso 2015; Al-Shehari 2017), the platform's technological interface (Alegria et al. 2013; Laxström, Giner, and Thottingal 2015), and the relationship between translation and technology in the encyclopaedia (O'Hagan 2016). For instance, Alonso (2015) found that professional translators often resorted to Wikipedia as an alternative to online dictionaries whenever they had to deal with

complex lexical terms. Similarly, Al-Shehari (2017) explored the feasibility and productivity of using Wikipedia in English-to-Arabic translation training courses. Alegria et al. (2013) analysed the interaction between MT and translators. In their investigation on the Basque Wikipedia, they closely monitored automated translations and post-editing[2] processes to improve the quality of MT-generated outputs. Following the implementation of Wikipedia's WYSIWYG[3] CAT Content Translation Tool (CX), the focus has shifted to developing and enhancing the device based on the feedback received by Wikipedia editors (Laxström, Giner, and Thottingal 2015).

During the last six years, more research has been conducted on Wikipedia translation. McDonough Dolmaya (2015), for instance, investigated the quality of a specific subset of translations in the English Wikipedia. By applying Mossop's taxonomy to the analysis of a sub-corpus of Wikipedia articles needing translation into English, she discovered that a significant number of transfer and grammatical errors persisted over time. Her findings revealed that such mistakes are likely to be corrected at a slow pace, with transfer errors taking more time to be amended. Expanding on some of these findings, O'Hagan (2016) observes that Wikipedia translation operates by a principle of 'self-repair', where articles created by one single user are subsequently and inexorably amended by their peers. She also posits that translators in the encyclopaedia have a high degree of autonomy because they can choose which articles they want to render into their language. Despite this freedom, however, O'Hagan (2016) contends that Wikipedia translation is far from a democratic 'open-to-all' practice, since it requires editors to be conversant with wikicode (the format or syntax used by wikis) and technology.

In a follow-up study, McDonough Dolmaya (2017) noted that Wikipedia lacks a specific translation policy. Cognizant of the knowledge gap between major and minority languages, she called for an official policy to avoid culturally biased translations into smaller Wikipedia communities. Nevertheless, she conceded that the solution is far from simple and remarked that more research into 'less structured' translation policies is necessary to gain better insight into Wikipedia's internal

[2] In Wikipedia, post-editing or 'edit' refers to any change that a page undergoes after the first publication.
[3] WYSIWYG stands for 'What You See Is What You Get'.

processes. More recent research has built on McDonough Dolmaya's work. For example, Shuttleworth (2017) proposed a series of foci to elicit data and further investigate translation phenomena in Wikipedia. He questioned previous assumptions of the collaborative aspect of Wikipedia translation, arguing that some editors 'are likely to be acting under their own steam and in isolation' (Shuttleworth 2017, 311). He refers to Wikipedia translators as 'editor-translators' and believes that their activities should be examined alongside other editing tasks since they both form a continuum.

In subsequent work, Shuttleworth (2018) examined the linguistic point of view (LPOV) expressed in a subset of translated Wikipedia entries describing the murder of the Russian politician and Putin critic Boris Nemtsov. He noticed that translated material in other Wikipedia communities tended to follow the LPOV of the original Russian text. Shuttleworth (2018) attributes this intended similarity to post-editing work, hence reinforcing his previous claims that translation and editing in the encyclopaedia are two sides of the same coin. Moreover, he acknowledges that post-editing can lead to the proliferation of hidden or undeclared translations. Shuttleworth (2018, 234) describes this phenomenon as the 'dark matter' of Wikipedia and posits that the obscure provenance of these texts is a valuable indicator of the changes undergone by a translated article over the years.

Along the same lines, Jones (2018b) provides a critique of McDonough Dolmaya's distinction between translation and original writing or editing. In his case study of the English Wikipedia translation of the entry for Paris, Jones (2018b) holds that translators in the user-driven encyclopaedia are not merely altruistic information bridges. By investigating the debate unfolding on the talk page attached to the entry, he observed that negotiations among Wikipedia editors are conflict-ridden and have an intersubjective tone. In a similar investigation, Jones (2018a) concluded that negotiations in Wikipedia are often multifaceted and that translator-editors 'compete as much as they cooperate'.

Finally, Torres-Simón (2019) also approached talk pages as foci of negotiation in Wikipedia. She studied lay understandings of translation across 93 language communities of the encyclopaedia. Her analysis of comments posted on talk pages delved into the most frequently discussed translation-related issues. Upon

comparing the various language versions of the Wikipedia entry for ‘translation’, Torres-Simón (2019) found a widespread tendency to portray translation as a classical activity, paying little, if any, attention to translation technologies and current trends. Nonetheless, she noted that some of the largest Wikipedia communities had reconciled, to some extent, those views with more contemporary attitudes to translation.

### 1.1.2 Wikipedia and automation

As stated in 1.1, there are relatively few studies on the use of automation in Wikipedia. This lack of research is revealing, given that the encyclopaedia relies upon a series of automated devices such as bots for the performance of crucial maintenance tasks. Geiger (2011) notes that, until the last decade, investigations into Wikipedia were typically driven by human factors, neglecting the relevance of automated software agents. The weight of automation in the online encyclopaedia is also attested by Halfaker and Riedl (2012), who argue that since their inception in 2002, bots have become an essential part of Wikipedia’s ecosystem and, ‘through their interaction with humans, they changed [the encyclopaedia’s] culture’.

The English Wikipedia defines bots as ‘automated tool[s] that carr[y] out repetitive and mundane tasks’ (‘Wikipedia:Bots’ 2021) that help maintain the platform. Such tasks include, but are not limited to, fighting vandalism, welcoming new users and sending warnings to disruptive ones, detecting and amending spelling errors, and finding articles and contributions that are likely to infringe Wikipedia’s policies. Besides, as will be explained in more detail in 2.4.1, bots have been used to create new content in several language communities. Clément and Guitton (2015) classify Wikipedia bots into two broad categories, ‘servant bots’ and ‘policing bots’. As their name suggests, servant bots are primarily invested in laborious maintenance work such as correcting orthography, adding templates to articles, and archiving inactive conversation threads (Lih 2009; O’Hagan 2016). On the other hand, policing bots are tasked with overseeing the enforcement of Wikipedia’s policies and undoing vandalism.

Access to bots in Wikipedia is restricted to a select group of senior editors with expertise in coding (Geiger 2011; de Laat 2015; Tsvetkova et al. 2017). Geiger

(2011) maintains that restrictions are intended to prevent the misuse of the device and minimise the risks that may arise from a malfunctioning bot. Despite careful policing, he acknowledges that errors and misinterpretations occurred at least until the beginning of the last decade. According to Geiger (2011, 91), early attitudes towards bots in Wikipedia were mixed, as some users viewed these devices as 'ruthlessly moral' and capable of 'taking a particular view of Wikipedia to its logical extreme'. Halfaker and Riedl (2012) also addressed the scepticism expressed by some users at the time, positing that bots '[could] be massively disruptive to the [Wikipedia] community if they perform[ed] inappropriate actions'.

De Laat (2015) has further investigated the interaction between bots and humans in Wikipedia. He offers a critical view of the vandalism patrolling system, which he regards as 'opaque'. He contends that policing or patrolling bots in Wikipedia often target non-registered users. Such users are automatically labelled as untrustworthy by the algorithms. De Laat (2015) views this approach as problematic and adds that bots 'may become a nuisance for good-faith contributors'. Notwithstanding these claims, research by Clément and Guitton (2015) has shown more optimistic results. Their study found that although policing bots were perceived negatively by some, they were generally regarded as 'indentured collaborators and potentially valuable helpers'.

Tsvetkova et al. (2017) maintain that bots in Wikipedia are unpredictable and that their actions are sometimes as inefficient as those performed by humans. This assertion has been contested by Geiger and Halfaker (2017), who argue that Tsvetkova et al. (2017) misinterpreted productive work as instances of conflict. Regardless of where one stands in the bot debate, it is worth mentioning that the studies discussed in this section have concentrated exclusively on bots as maintenance and policing devices. Clément and Guitton's (2015) categorisation, while illustrative of how Wikipedia operates, does not account for bots whose primary function is to fill knowledge gaps.

Bots have been used in several Wikipedia communities to create short articles on a wide range of topics, from landmarks and short biographies to animal and plant species. Although bots cannot be considered translation devices, their primary function is to assist in the creation of multilingual content. They do so by importing

data from a large number of machine-readable databases and feeding that information into templates that are later employed to generate articles across various Wikipedia language communities. For example, Lsjbot, one of the encyclopaedia's most popular and notoriously controversial bots (Jervell 2014), gathered data from multiple sources and created 9.5 million entries in the Swedish, Cebuano, Dutch and Waray-Waray language versions.

The device, developed by the physicist Sverker Johansson, remained active in the Swedish Wikipedia for more than three years until its usage was interrupted in 2016 following complaints of 'lack of meaningful content and human touch' ('Lsjbot' 2021). Johansson responded harshly to this criticism by highlighting Wikipedia's inherent cultural and gender biases. In a 2014 interview, he said that if Lsjbot had not created so many articles, they would have been written chiefly 'by young, white, male nerds and reflect male interests' (*The Local* 2014). Controversy aside, the last statistical report released by Wikimedia Foundation in January 2019 shows that 81% of the Swedish Wikipedia entries and 54% of the Dutch Wikipedia entries were bot-generated[4] ('Wikipedia Statistics' 2019).

The lack of research on bot-created content shows some parallelisms with the little attention that CX has received since its inception in 2014. Although CX has been running for only seven years, the tool statistics indicate that CX has had a considerable impact on Wikipedia translators. There are over 958,000 articles generated using CX at the time of writing, spanning 200 language communities ('Content Translation Statistics' 2021). The English Wikipedia remains the principal source of the translations. The Spanish, French and Arabic Wikipedia communities top the ranking of CX-generated articles. The Spanish Wikipedia hosts ca. 92,000 CX-generated entries, the French Wikipedia is home to ca. 73,000 entries, and the Arabic Wikipedia contains approximately 52,000 entries ('Content Translation Statistics' 2021).[5]

Even though 958,000 may not come across as an exceptionally high number in an encyclopaedia that comprises 56 million entries, CX's usage has soared in the past

[4] The data only include Wikipedia articles that were first created using bots. Consequently, modifications (post-edits) made by either bots or human contributors are not part of the count.
[5] Likewise, the CX stats exclusively refer to Wikipedia articles generated with the device.

lustrum. McDonough Dolmaya (2017), drawing on data from December 2016, situated the number of CX-generated articles at 94,210. She also observed that English was the source language for 60% of the translations. Since then, the use of English as a source language has also increased among Wikipedia users of the tool, accounting for nearly 70% of the translations as of 2021 ('Content Translation Statistics' 2021). Moreover, CX stats reveal that 65,023 registered Wikipedia users have translated from English into their languages using the device in the past five years. Of these, 13,995 are Spanish Wikipedia editors, and 10,586 are registered in the French Wikipedia ('Content Translation Statistics' 2021).

Despite the growth in CX usage since McDonough Dolmaya's (2017) study, the limited research to date can be narrowed down not to numbers but to the affordances of the tool. Laxström, Giner, and Thottingal (2015), who were heavily invested in the development of CX, surveyed 106 English Wikipedia users of the program to gather information about its reception. Their findings show that, while most respondents were concerned with the quality of the translations, a substantial majority believed that using CX could make the process smoother (Laxström, Giner, and Thottingal 2015).

## 1.2 Research design

Thus far, most studies have emphasised the collaborative and sometimes disruptive nature of Wikipedia translation. Nevertheless, little research has been undertaken on volunteer translation as practice and the mechanisms that underpin its performance. As pioneering as McDonough Dolmaya (2012) study was, her research relied upon anonymous surveys and targeted editors of one Wikipedia community. The vast majority of previous studies have also overlooked the role and functionality of translation guidelines in the user-generated encyclopaedia (McDonough Dolmaya 2017). To date, most research on Wikipedia has concentrated on translation quality, the intricate processes of knowledge building and dissemination, and understanding the editors' motivations for contributing to Wikipedia.

Moreover, the previous section showed that there is a paucity of studies on the use of content-creation devices in Wikipedia. Most research on bots has tackled the role

of these automated devices in maintenance and policy enforcement, but bot-generated content in multilingual contexts remains understudied. The same applies to CX, despite its usage having grown exponentially in the past years. McDonough Dolmaya and Sánchez-Ramos (2019) draw attention to the study of automation, observing that further research is necessary to determine, among other things, whether 'the use of these tools as part of the translation or authoring workflow affects the motivation of volunteers'.

Consequently, this thesis seeks to examine the normative aspects and the material arrangements that underpin the practices of Wikipedia translators of four language communities. To this end, the investigation will draw upon practice-theoretical concepts, particularly on Wenger's (1998) 'communities of practice'. Despite discrepancies over using the lay term 'community' to refer to Wikipedia (Ensslin 2011), the concept has been employed in praxeological studies on Wikipedia's hierarchical structure (O'Sullivan 2009; Usman and Yennita 2018). As Chapter 2 will explain in more detail, framing the study of individual language versions of Wikipedia around the concept of communities of practice can contribute to a better understanding of how editors negotiate, approve and assimilate local translation standards. For instance, one central principle underlying communities of practice is that distinct individuals engage with one another to fulfil a set of specific goals or 'joint enterprise'. To achieve those aims, community members have access to a shared repertoire consisting of common values, concepts and material dispositions that are unknown to non-members or outsiders (Wenger 1998).

Building on previous studies on multilingualism in Wikipedia, this investigation will tackle the Spanish, French, Dutch and Swedish language communities. The selection of these communities is motivated by the underlying differences in the use of automation among them, as addressed in 1.1.2. In particular, this study aims to examine the impact that CX and other cross-lingual content-creation devices such as bots have had on the translators' practices in those Wikipedia communities known for their deployment of automation. Furthermore, as previously stated, comparative research on Wikipedia has highlighted the relative autonomy of its language communities, where volunteers have leeway to negotiate standards and restrict access to certain materials. On this premise, this thesis also seeks to

ascertain whether local Wikipedia translation standards influence the editors' activities.

To achieve this, the research design was engineered around three objectives. The first aim is to investigate how translation standards are regulated and negotiated across the four language communities of Wikipedia. The second is to elucidate whether, and if so, how, experienced translators of those communities have incorporated those local standards into their practices. Thirdly, this study sets out to analyse how and to what extent automated devices such as CX and bots have influenced the participants' practices and contributed to their evolution, focusing on the last lustrum. To meet these research aims, the investigation will draw on data gathered from a series of Wikipedia documents (policies, guidelines and essays) on translation, their ancillary talk pages, and semi-structured interviews with 16 experienced translators of the encyclopaedia. Section 1.3 below and Chapter 3 will provide further details on data collection and analysis.

Overall, this research contributes to enhancing knowledge on the role of documented standards and automated devices in Wikipedia translation. It also extends prior studies on the impact of automated devices on content creation in Wikipedia by analysing the translators' views on and usage of CX and bots across communities. Having first-hand interviews with volunteer translators about their work in Wikipedia may also help develop a better understanding of the intricate processes of peer co-production. Ultimately, ascertaining the relevance of standards and automated devices for translation may prove beneficial for Wikipedia communities as a whole. The findings could a) shed light on how the negotiation of translation guidelines transpires, b) raise awareness about problems stemming from the use of automation, and c) stimulate the development of new policies and devices to tackle unresolved issues.

## 1.3 Research Questions

This thesis explores the following overarching research question:

> How have *regulation*, *negotiation* and *automation* contributed to the emergence and evolution of translation practices across the Spanish, French, Dutch and Swedish language communities of Wikipedia over the last five years?

To better understand how regulation, negotiation and automation work towards configuring translation practices in the user-driven encyclopaedia, the ensuing specific research questions have been proposed:

1. How and to what extent have the four Wikipedia language communities regulated translation practices?

This study aims to address this question by conducting a thematic analysis of documented standards of practice with data gathered from Wikipedia translation guidelines/essays and bot-creation policies when available. Specifically, the data will be retrieved from all the existing Wikipedia pages tackling translation in the four communities under investigation. The first part of the analysis will look at the most salient features of the documents to gain insight into those aspects that are most relevant to each community. The themes emerging from this examination will then be compared and categorised following two criteria: universality (shared by all four communities) and locality (specific to one or more communities).

2. How are translation standards negotiated in each of the four Wikipedia language communities?

This question was devised to examine how editors engage with one another and negotiate changes to the documented standards analysed above. Data from the ancillary Wikipedia talk pages are collected and codified following a thematic approach. Those aspects that require clarification or engender controversy are scrutinised. The analysis will also resort to revision histories to ascertain whether some discussions and changes to the standards are correlated.

3. To what extent have experienced Wikipedia translators incorporated the standards set by their communities into their practices?

After probing into the standards and their negotiation, this question investigates whether experienced Wikipedia translators have implemented any of their provisions. To this end, the analysis focuses on data elicited from semi-structured interviews with 16 participants, four per language community. The interviews will tackle, among other things, the participants' knowledge of translation standards and other Wikipedia policies.

4. How and to what extent have automation and metadata contributed to changes in translation practices in Wikipedia over the last five years?

The final part of the study draws on the same interview dataset but delves into the participants' experience with and views on deploying automated devices such as CX and bots. Drawing primarily upon Shove's (2017) postulate that materials configure practices, Chapter 6 is aimed at ascertaining the impact of both internal and external devices on the evolution of translation in Wikipedia.

## 1.4 Overview of the thesis structure

This thesis comprises seven chapters, including the present introduction, the theoretical framework, a chapter devoted to explaining the datasets and methods, three analytical chapters, and the conclusion. This first chapter contextualises the research and outlines the aims, scope and structure of the study.

Building on previous studies, Chapter 2 introduces the theoretical foundations that sustain the thesis. Section 2.1 begins by discussing different understandings of practice; section 2.2 concentrates on Wenger's (1998) concept of 'communities of practice'. Drawing mainly on Warde (2005; 2016), section 2.3 revolves around the normative aspect of practice. Section 2.4 tackles the materiality and mutability of practice. This chapter further elaborates on how those theoretical tenets could be applied to a praxeological study of translation in Wikipedia.

Chapter 3 presents an overview of the methods and datasets that will inform the investigation. The chapter describes three differentiated, albeit complementary, analytical stages: a) Documented standards, focusing on translation guidelines/essays and bot-generated content policies; b) their negotiation on the ancillary talk pages or discussion forums; and c) semi-structured interviews with

16 experienced Wikipedia translators. The chapter also explains the data selection criteria and reflects on the methodological challenges and limitations of the research.

The first of the three analytical chapters, Chapter 4 examines the extent to which the four Wikipedia communities have regulated translation practices (RQ1) and how documented standards are negotiated in each community (RQ2). It begins by analysing documented standards on translation and the use of automation, first looking for commonalities across the four Wikipedia communities and then focusing on the differences across the pages. The second part of the chapter shifts the focus to the editors' negotiation of both translation and bot-related issues on the talk pages. In particular, it investigates the most salient aspects of each document with data retrieved from postings.

Having examined the documented standards of practice, Chapter 5 investigates how senior Wikipedia translators have incorporated them into their practice (RQ3). Section 5.1 introduces the participants of the study. Sections 5.2 and 5.3 delve into their background, particularly their motivations to become translators, and the challenges they encountered during their socialisation into the practice in Wikipedia. Sections 5.4 and 5.5 concentrate on the participants' knowledge and subsequent incorporation of Wikipedia's standards.

Still relying primarily on interview data, Chapter 6 analyses the participants' use of and views on automated devices such as CX and bots. The chapter pays attention to how this software has configured translation practices across the last lustrum (RQ4) and forecasts what may change in the years ahead based on the participants' answers. The final part of Chapter 6 concentrates on the potential role of online repositories such as Wikidata in configuring translation in the user-driven encyclopaedia.

The thesis concludes with a summary of the most revealing findings and discusses how the investigation has contributed to gaining a better understanding of the documented standards and devices that underpin the practices of a selected group of Wikipedia translators. The last section of the chapter offers some concluding remarks and provides suggestions for further research.

# Chapter 2. Conceptualising Wikipedia's translation practices

This chapter sets out to introduce Wenger's (1998) concept of 'community of practice' as a suitable framework to better understand how Wikipedia translators engage with one another and negotiate changes to the documented standards approved by their language communities. Drawing on an understanding of practice as 'the property of a community created over time by the sustained pursuit of a joint enterprise' (Wenger 1998, 45), this project first and foremost places emphasis on the social or collective value of practices (Lave and Wenger 1991; Wenger 1998; Buch and Schatzki 2019), but also on the significance of materials for the realisation, perpetuation or discontinuation of specific practices (Nicolini 2012; Shove, Pantzar, and Watson 2012; Shove 2017).

To this end, each section of this chapter explores various, albeit not exclusive, strands of practice theory, and their potential application to the study of translation in Wikipedia. Section 2.1 provides an overview of practice theory and different understandings of practice. Section 2.2 then presents communities of practice as a type of praxeological thinking or practice-oriented approach. Section 2.3 shifts the focus to conventions or standards of performance as an underlying feature of practices. Finally, section 2.4 discusses the material dimension of practice, which is of utmost relevance to comprehend how devices contribute to the configuration of practices and the evolution of 'old ways of doing' (Reckwitz 2002).

## 2.1 Introducing practice theory

Most prominent practice theorists agree on the fact that there is neither a universal definition of practice nor a unified practice theory (Nicolini 2012, 8; Gherardi 2017, 38; Olohan 2017, 161; Schatzki 2018, 153). For instance, in his seminal paper, Reckwitz (2002) provides a comprehensive definition of practice that contrasts with Wenger's (1998) more goal-oriented approach. According to Reckwitz (2002, 249), practice 'consists of several elements, interconnected to one other: forms of bodily activities, forms of mental activities, "things" and their use, background knowledge

in the form of understanding, know-how, states of emotion and motivational knowledge'.

Likewise, Schatzki (2010, 129) refers to practice as 'an evolving domain of varied activities linked by common and orchestrated understandings, rules and normative teleologies'. Thus, practice is not a repetition of unconnected actions but, rather, a set of interrelated conventional activities and processes external to the individuals that, when put together, contribute to the successful performance of well-defined entities. For instance, to play tennis, one should utilise the right equipment, follow certain conventions and have a basic knowledge of the sport. This codified knowledge is nothing more than a feature or quality of the practice itself; the individuals or practitioners are perceived as mere 'carriers' of practices (Reckwitz 2002, 250; Schatzki 2003, 182; Shove, Pantzar, and Watson 2012, 70; Olohan 2017, 162).

Although neither Reckwitz (2002) nor Schatzki (2010) view practice as an activity that results from a strict communal desire to achieve specific objectives, they both consider practice to be a social endeavour. Moreover, as will become apparent in 2.2, some core elements in Reckwitz's and Schatzki's conceptualisations are also present in Wenger's (1998) interpretation of practice. In particular, Wenger acknowledges that to perform practice, it is essential to have access to artefacts ('things'), common values, symbols, and documents. Another point of convergence is found in Wenger's (1998, 90) postulate that practices may, and often do, evolve over time because new practitioners ('newcomers') are likely to be more innovative performers than their senior counterparts. To better understand these differences in praxeological thinking, it is necessary to take into consideration their historical context.

The foundations of practice theory were developed between the 1970s and the 1980s. The early practice theorists drew some of their tenets from the works of Bourdieu, Giddens, Wittgenstein, Heidegger and, to a lesser degree, from the philosophy of Marx and Foucault (Reckwitz 2002, 243–44). However, as noted by Nicolini (2012, 25), the roots can be traced back to Ancient Greece. Aristotle, inspired by Plato's *The Republic*, distinguished between three types of knowledge: *phronesis* or 'practical wisdom', *episteme* or 'scientific knowledge', and *techne* or

'instrumental rationality'. Within this division, the Greek philosopher situated *praxis* (practice) outside the system of universal rules and thus beyond the realm of *episteme* or scientific enquiry (Aristotle 2001, 72). Therefore, while praxis was necessary for the development of humankind, it had to be avoided by philosophers because of its inherent mutability and materiality. Aristotle, in turn, referred to materiality or the production of artefacts as *poiesis*, which he considered to be part of *techne* or craftsmanship (Nicolini 2012, 26–27).

After a relatively long period of neglect, a renewed interest in practice emerged in the 19th century with the publication of Marx's *Das Kapital* (1867) and, most notably, *Theses on Feuerbach* (1888). Although Marx never employed the term 'practice', he believed that praxis was essentially any activity aimed at changing the world (Nicolini 2012, 31). Core to Marxian thought is the idea that practice is as important as theory and that by engaging oneself in practical matters, one is equipped to transform society (Marx 1962, 397). Drawing on some of these Marxian principles, Heidegger (1929) took the study of practices into the field of linguistics. For Heidegger, language practices determine how humans perceive the world. He also argued that humans think in practical terms most of the time, except when they experience disruption, in which case they theorise in an attempt to find a solution (Nicolini 2012, 35). Within this process of theorisation, Wittgenstein (1981) posits the idea that meaning is socially constructed by engaged agents.

According to Nicolini (2012, 47) and Warde (2016, 35), Giddens, for whom practices are mere regularised types of acts or activities, emphasises the practitioners' willingness to engage in a task as one of the driving forces responsible for both the evolution and the perpetuation of certain practices. Similarly, Bourdieu (1977) recognises the relevance of participation in the study of practice (Nicolini 2012, 53). Central to everyday life, practices are constrained by an external social force known as habitus, which Bourdieu defines as 'principles of the generation and structuring of practices and representations which can be objectively regulated and regular without in any way being the product of obedience to rules' (Bourdieu 1977, 73). In essence, habitus is the property of a group: it is what makes people react objectively in a given context, usually in anticipation of a future event. While habitus is inherently social, it is by no means intentional: its strategic power stems from the capacity of two individuals to anticipate each other's actions and respond

accordingly through the performance of practices. In Bourdieu's words, 'habitus is the source of these series of moves [practices] which are objectively organised as strategies without being the product of a genuine strategic intention' (Bourdieu 1977, 73). Although modern practice theory has not embraced Bourdieu's concept of habitus and has tended to focus on praxis or the performance of practices, his work paved the way for the development of this theoretical domain (Nicolini 2012, 44).

From what has been discussed so far, it is clear that practices are intrinsically social. As Schatzki (2018, 153) notes, practices play a crucial role in social affairs and thus define the way people interact with the world; they are ubiquitous and form constellations. More recently, Schatzki (2018) referred to constellations as a series of loosely connected practices or bundles that go beyond the entity they integrate. In other words, certain practices can share similar or even the same materials. For example, dictionaries have long been used in translation (Olohan 2021), but they also configure the practices of linguists, lexicographers and teachers. Section 2.2 will return to the issue of constellations from a community of practice perspective (Wenger 1998).

One core aspect of practices, aside from their social character, is that they are composed of both explicit and implicit elements. Explicit components are as influential for the performance of practices as the implicit ones. They comprise the environment or space where practice takes place as well as the materials without which the realisation of certain practices would be hampered or even impossible. Wenger (1998, 83) acknowledges the importance of materials for the completion of practice, alongside implicit elements such as routines, gestures, symbols, and concepts. He utilises the term 'shared repertoire' to describe the interrelation of values, actions, concepts, and materials in practice.

Like Wenger (1998), Schatzki (2018, 154) supports the view that practices are entwined with materiality. Nicolini (2012, 171) postulates that it is not possible to approach practices without considering the materials that contribute to their performance. In a similar vein, Shove, Pantzar, and Watson (2012, 82) hold that practice as a unit of analysis involves an amalgamation of diverse meanings, materials and competence. By way of illustration, the successful performance of a

given practice such as translation would not be possible without the conceptual (skills, expertise) and material arrangements (computer, CAT tools, etc.) required to complete the tasks.

Although the significance of materials for the emergence and evolution of practices will be discussed in more detail in 2.4, it is worth noting at this stage that materiality plays a fundamental role in the survival of practices. In brief, if a material is no longer available, the practice that made use of it can either vanish or evolve into something else. Conversely, as practices change over time, some materials become obsolete. This interdependence between materials and practices has led some scholars such as Shove, Pantzar, and Watson (2012, 52) to claim that practices depend primarily on materials for their development and persistence. Wenger (1998, 77) seems to agree that practices are dynamic and that it is up to the practitioners to modify and adapt them to suit their needs if there is communal consensus.

Alongside materiality, the dynamic nature of practices means that they can adapt quickly to meet the current demands of the practitioners. Throughout history, practices and the materials that configure them have evolved; their past, usually documented and stored in books, manuals or repositories, allows practitioners to look in retrospect, learn from trial and error, and benefit from experience. For example, a few practices that are nowadays assumed to be fundamental for the successful performance of tennis precede the sport, and some can even be traced back as far as 12$^{th}$-century France. Gillmeister (1997, 117) observes that historical records report the practice of striking a ball with the palm in medieval France.

Documented experience and sustained updates in technology have played a significant role in the evolution of concrete practices and the discontinuation of others. For instance, in their analysis of the history of driving, Shove, Pantzar, and Watson (2012, 40) point out that the early practices of driving a car were inspired by horse riding and cycling. This idea of inspiration also applies to user-driven projects such as Wikipedia, where practices like typing, writing encyclopaedic articles, and exchanging opinions in forums were already established before the platform was created in 2001. Moreover, translation has a long history as a practice performed by volunteers rather than professional translators. Thus, it can be argued that both editing and translation in Wikipedia are connected to one another and to

other previous practices. The following subsection provides an overview of some of the most relevant practices in the online encyclopaedia.

### 2.1.1 Practices and practitioners in Wikipedia

Each Wikipedia community comprises a diverse set of practices and practitioners, from users and editors to project administrators and other functionaries.[6] Wikipedia users or 'Wikipedians' are those individuals with a registered account in the project. Within this group, Wikipedia distinguishes between active and passive users. To be considered an active member of the community in a given month, 'one or more actions [contributions] need to be made' during the said month ('Wikipedia' 2021). Active, engaged users are often referred to as 'editors'.

According to Wikipedia, editors 'are the volunteers that write or edit Wikipedia's articles, unlike readers who simply read them' ('Wikipedia:Wikipedians' 2021). This category comprises editors, including translators and administrators. Project administrators or system operators (sysops) are senior editors and trusted members elected by their communities with access to specific tools that allow them to delete articles that do not follow Wikipedia's policies, block disruptive users, remove inappropriate content, and – in the case of 'bureaucrats' (a subtype of functionaries) – rename users' accounts.

Wikipedia is a relatively flexible project in which volunteers can take on diverse roles (O'Sullivan 2009). As explained in 1.1, despite this flexibility, editors abide by a series of standards, some of which are approved by their local language communities. As a consequence of these regulations, practices such as editing in the user-driven encyclopaedia are known to be influenced by established standards. For example, when writing a Wikipedia article, it is necessary to comply with a set of overarching policies that go from using a neutral, objective style (WP: NPOV) to including reliable sources (WP: VER). There are also more technical aspects or know-how that are part of the editing process; these include writing in wikicode,

[6] Functionaries in Wikipedia are usually trusted, senior editors that have been elected by their local communities to perform maintenance tasks that require special permission and the use of restricted tools. Some of these tasks include renaming user accounts and tracking fraudulent users with multiple accounts (sock puppets).

adding internal links to other articles, and assigning an article to one or more categories.[7]

Another activity that makes up the practice of editing is categorisation. Wikipedia articles are categorised to make it easier for readers to find information about a topic of their interest. For instance, the English Wikipedia article about the former British Prime Minister Theresa May can be found in over 30 categories which the reader can access from the bottom of the page ('Theresa May' 2021). Some of these categories provide information about her age (1956 births), gender and nationality (20th-century British women, 21st-century British women), her birthplace (People from Eastbourne), and her political affiliation (Leaders of the Conservative Party, UK), among other qualities.

If editing plays a significant role in making content more accessible to readers, so does translation. Wikipedia articles or entries can be, and often are, translated from one language into another. This can be done manually or with the aid of devices such as CX. The output is always an independent entry, which can change over time as more editors modify content in the target language (O'Hagan 2016; Shuttleworth 2018). Besides, translators in Wikipedia can choose what part of the entry they want to translate, often omitting or expanding sections of the text. The English Wikipedia help page entitled 'Translation' states that 'articles on a given subject in different languages are typically edited independently, and need not correspond closely in form, style and content' ('Help:Translation' 2021). Moreover, according to the guidelines, 'if portions of an article appear to be low-quality or unverifiable, [editors can use their] judgment and do not translate those portions'.

From what has been discussed so far, it is apparent that translation in Wikipedia allows for some flexibility and relies on a combination of personal judgment, observance of certain rules and implicit knowledge. However, it is not yet clear whether translation practices in the online encyclopaedia are regulated by specific standards other than those that underpin editing. Consequently, as explained in the previous chapter, this thesis sets out to investigate the recommendations given by local translation standards in four language communities of Wikipedia, how they

[7] Categories in Wikipedia are aimed at grouping together articles that share one or more common features such as 'place of birth', 'profession' or 'nationality'.

are negotiated, and the extent to which experienced translators have incorporated them into their practice. The next section introduces Wenger's (1998) concept of 'communities of practice' as a type of praxeological thinking that places emphasis on the practitioners rather than on practice as a unit of analysis.

## 2.2 Communities of practice

As happens with the term 'practice', the theoretical concept of communities of practice or communities of practitioners is more obscure than it might seem at a superficial level. This conceptualisation appears in the work of some practice theorists, most notably in Lave and Wenger (1991) and in the book published by the latter entitled *Communities of Practice* (1998). However, as noted by Nicolini (2012, 77), the underlying principles that sustain this line of thought go as far as Ancient Greece and, in modern times, they are present in the works of Durkheim, Weber, and Mauss.

The term 'community', which entails socialisation, has been attached to the label 'practice' on the grounds that practices cannot transpire outside social settings (Wenger 1998; Nicolini 2012, 78). Communities of practice can emerge even when physical distance prevents practitioners from having a closer interaction. As the outbreak of the Covid-19 pandemic and subsequent quarantine measures proved in 2020 and 2021, certain communities of practice (schools, universities, companies) had no other option but to resort to remote working for a prolonged period. Besides, as discussed in Chapter 1, research on online communities of practice is not a recent phenomenon (O'Sullivan 2009; Wenger, White, and Smith 2009; Townsend, Osmond, and Phillips 2013).

Wenger (1998, 72) posits that not all communities are defined by practice. He illustrates this point with the example of residential neighbourhoods, which are commonly referred to as communities, even though dwellers share little more than their geographical location. Therefore, critical to Wenger's definition of community of practice is the existence of a joint enterprise, common goal or domain, which necessarily implicates mutual engagement among practitioners around a shared repertoire (Farnsworth, Kleanthous, and Wenger-Trayner 2016, 143). This repertoire consists of material elements such as artefacts and tools, but also styles,

actions and concepts (Wenger 1998, 72–73). For individuals to become practitioners, they must engage in the same activities, and they should all share one or more goals – a joint enterprise – towards which all their actions are directed (Wenger 1998, 95). Thus, engagement involves the participation of competent individuals willing to endorse the standards and values of their community of practice (Wenger 1998, 76).

Practitioners of a given community may, and often have different backgrounds and ideas. For Wenger (1998, 76), however, mutual engagement supersedes any differences that practitioners might have to the extent that their 'identities become interlocked'. These links do not mean that practitioners lose their identities but, instead, that all their attempts should be directed towards the successful completion of practices. As part of this process, Wenger (1998) contends that disagreement, challenges and competition among members of a community should be regarded as signs of commitment, even more so than 'passive conformity'. He also observes that 'rebellion' – contesting elements of the shared repertoire – has the potential to be a driving force of change (Wenger 1998, 77). As such, it could be necessary for the evolution of some practices over time.

Communities of practice have hierarchies and boundaries, which means that not all practitioners have the same status or position within the community. According to Wenger (1998, 113), participants within a given community 'develop idiosyncratic ways of engaging with one another, which outsiders cannot easily enter'. He employs the term 'outsiders' to refer to individuals who do not have access to a shared repertoire, and therefore are not familiar with the intricacies of the enterprise that brings all members of a community together (Wenger 1998, 114). He also utilises the term 'periphery' to describe those instances in which outsiders are occasionally allowed access to a specific community with the aim of learning from more experienced members (Wenger 1998, 120).

For Wenger (1998, 120), peripheral participation is an ambiguous arena, a limbo in which 'access to a practice is possible, but it can also be a position where outsiders are kept from moving inbound'. Successful individuals in the periphery can, however, end up becoming full members of the community (Farnsworth, Kleanthous, and Wenger-Trayner 2016, 155). Shove, Pantzar, and Watson (2012,

72) refer to experienced members as ‘full practitioners’ and place them at the ‘core’ of their communities, with ‘novices’ remaining in the periphery. Membership, whether full or peripheral, is determined by practice rather than by the specific qualities of an individual. This idea is shared by Nicolini (2012, 139), who contends that ‘activity performs membership’.

Another term Wenger (1998, 154) uses to denote the process through which outsiders can join a community to achieve full membership is ‘inbound trajectory’. Shove, Pantzar, and Watson (2012, 70) also employ similar terminology when they describe the assimilation of a given practice as ‘career progression’. Wenger (1998) and Shove, Pantzar, and Watson (2012, 70) agree on the significance of learning for becoming a full-fledged practitioner, with the latter holding that outsiders or novices ‘become members of specific communities of practice through learning and experience’. To better understand this process, as stated in 1.4, chapters 4 and 5 of this thesis will examine the function of Wikipedia translation standards and their incorporation by a selected group of experienced translators.

Shove, Pantzar, and Watson (2012, 70) contend that, during the learning process, there is a major turning point in which newcomers start seeing themselves differently. The time required to become a full or full-fledged practitioner depends on the nature of each community, the degree of exposure to the practice, and the amount of input that novices receive from both experienced members and the environment. Shove, Pantzar, and Watson (2012, 72) consider that the learning process, the career path, is far from being unidirectional: full-fledged practitioners often benefit from novices, who may propose changes to established rules or expectations. This idea is shared by Lave and Wenger (1991, 94), for whom ‘mastery resides not in the master but in the organisation of the community of practice of which the master is a part’.

For Wenger (1998), being a member of one or more communities of practice entails participation. He differentiates between engagement and participation in practice, noting that the latter cannot be switched off. Thus, while engagement is closely related to performance, participation is a core component of the practitioners’ identity. For instance, in the case of Wikipedia, translating an article from one language to another and complying with the community’s shared repertoire can be

classified as engagement. Actively identifying oneself as a Wikipedia editor or 'Wikipedian' as a result of that engagement is, however, an example of participation. As a study primarily concerned with translation as performance, this thesis will concentrate on mutual engagement.

Regarding engagement, Wenger (1998) observes that, although it is a social endeavour, it does not necessarily involve interacting with other people. He notes that preparing oneself for a formal event such as an academic workshop or presentation shows a high degree of engagement in practice. As discussed in 1.1.1, this capacity to engage in practice without necessarily interacting with fellow practitioners has been documented in Wikipedia, where some translator-editors work in isolation, without always receiving feedback from their peers (Shuttleworth 2017, 311).

In subsequent work, Wenger, McDermott, and Snyder (2002) found that some communities of practice often suffer from what they called 'disorders', which is nothing more than the inability or unwillingness to accept change and welcome new ideas. In this environment, innovation and rebellion cannot flourish, leading to long-term stagnation. They identified three primary disorders: a) domain-related, b) community-related, and c) practice-related. Domain-related disorders occur when members of the community claim exclusive ownership of the joint enterprise, not allowing others to engage. One common type of domain-related disorder is imperialism, where members seek to impose their views at any expense. Other disorders in this area include narcissism, where members resort to the joint enterprise to push their personal agendas; marginality, where certain individuals are excluded from making decisions; and factionalism, where disagreement between members can become irreconcilable, leading to the emergence of clearly differentiated sides.

Concerning community-related disorders, Wenger, McDermott, and Snyder (2002) posit that sometimes members can create 'tight bonds', leading them 'to act in ways that would shock outsiders'. Because of these close interrelationships, communities can become cliquish (dominated by a core group of members), egalitarian (where standing out is penalised), and dependent (where members rely on a reduced group of leaders to make decisions for them). Furthermore, communities may also become

disconnected (where the boundaries are too vague), stratified (where a few members have power), and local (where a community fails to expand their boundaries).

The third type of disorder, affecting the realm of practice, stems from the shared repertoire. Wenger, McDermott, and Snyder (2002) observe that practice-related disorders occur when a community fails to update and extend their values, methods and tools. They regard this issue as particularly problematic, because failure to innovate and embrace change can lead some members to abandon their community. According to Wenger, McDermott, and Snyder (2002), the most frequent practice-related disorder is documentism, which happens when communities attach excessive importance to documentation. Both the accumulation of documents and an overreliance on written or documented standards can hinder progress and overwhelm practitioners. Chapter 4, and to a lesser extent Chapter 5, of this thesis will draw on Wenger, McDermott, and Synder's (2002) concept of 'disorders' to gain a better understanding of the problems that may arise when communities attempt to regulate and negotiate translation practices in Wikipedia.

Besides disorders, another issue with communities of practice is that they may lack clear boundaries. As noted by Schatzki (2018), one of the downsides of approaching communities of practice in terms of engagement, recruitment and joint enterprise is that any of these concepts could be extended to embrace larger organisations, cities or countries. Once more, perhaps one of the most illustrative examples can be found in Wikipedia. Volunteering in a large project such as Wikipedia shows a joint enterprise and a commitment to a shared repertoire. Yet, the encyclopaedia is linguistically and culturally diverse: it comprises 323 communities of contributors, each representing different languages and language varieties spoken around the world.

The issue of scope or outreach of the communities of practice is addressed by Wenger (1998, 128), who uses the concept of 'constellations' to refer to several communities of practice that seem to be loosely connected by either a joint enterprise or overlapping values. A few of these communities may have members in common who can occasionally act as 'brokers'. According to Wenger (1998, 108), brokering involves importing skills acquired in one community into another,

even when this is done unconsciously. In the case of Wikipedia, editors are not limited to one language community, even though their degree of engagement may vary considerably across projects.

It is worth noting that the term 'communities of practice' has raised criticism from other practice theorists, some of whom have questioned the suitability of describing practitioners as a 'community'. Nicolini (2012, 92), who revisits Wenger's communities of practice, holds that the term is ambiguous and slightly redundant given the implied sociality of practices. Practice, as mentioned earlier in this chapter, is, by definition, social (Nicolini 2012; Olohan 2021). Both the endorsement and accomplishment of a practice rely on more than one individual. Reckwitz (2002, 250) reinforces this point by claiming that 'to say that practices are 'social practices' then is indeed a tautology'.

Similarly, Hui (2017, 64) believes that Wenger 'draws too strongly upon positive framings of the term [community]'. Nicolini (2012, 92) raises the possibility of taking the word out of the equation and referring to communities of practice only as practices, since all the defining elements of communities of practice are those contained in their constituent parts. Despite this observation, he points to the impracticalities of pursuing such an enterprise, for a) the term is well established and has been endorsed by theorists (see O'Sullivan 2009), and b) it emphasises the connections between practitioners, who may identify themselves, cognitively or normatively, as being members of a community (Nicolini 2012, 93–94).

### 2.2.1 Communities of practice in Wikipedia

> New members are prospective contributors and are, therefore, Wikipedia's most valuable resource. We must treat newcomers with kindness and patience.
>
> Wikipedia: Please do not bite the newcomers

Before embarking on the study of translation in Wikipedia, it should be noted that a reduced number of scholars in the field have previously used Wenger's (1998) framework to examine the practices of translators. Neather (2012), for instance, draws on Wenger's (1998) conceptualisation to investigate the negotiation of meaning in close interactions between museum curators and translation agencies in China. His research found that both communities of practitioners engaged with one

another driven by a common interest or joint enterprise: the production of translated texts for museums. During the course of their mutual engagement, curators and translators redefined each other's identities. On the one hand, professional translators often required the assistance of curators to deal with technical vocabulary. On the other, curators depended on the translators' expertise to produce the texts. In light of this interdependence, Neather (2012, 266) concludes that none of the communities of practitioners could claim sole ownership of meaning, and that who counts as professional and in what context is a matter of perspective.

More recently, Yu (2019) also applied theoretical concepts from Wenger's (1998) framework to the analysis of online collaborative translation in Yeeyan, one of China's largest crowdsourcing platforms. Similar to Wikipedia, Yeeyan relies on the collaboration of thousands of volunteers interested in the dissemination of knowledge. However, unlike Wikipedia, Yeeyan is translation-oriented and its content targets a more select group of Chinese readers. In her study, Yu (2019) concentrates on issues of identity and negotiation of meaning among members of the platform. She describes Yeeyan as a large community of practice comprising various sub-communities where members can take on multiple roles.

To examine the members' commitment, Yu (2019) draws on Wenger's (1998) definition of participation, already discussed in 2.2, and reification. The latter can be described as the process through which members of a given community of practice give form to their experience (Wenger 1998, 58). As Yu (2019) notes, in a community such as Yeeyan, reification transpires in the translation of texts, but also in the processes of editing and decision-making. In brief, reification is intrinsically linked to experience: it denotes how practitioners perceive the world around them.

One notable aspect of Yu's (2019) study is that she replaces Wenger's (1998) concept of mutual engagement for that of 'mutual recognition' to better reflect the dynamics of the platform, where members not only engage with one another but also acknowledge and praise each other's work. After interacting with what she refers to as 'Yeeyaners' and observing their work for several months, Yu (2019) found evidence that suggests that the translators' offline and online experiences were often at odds with one another. This 'interference', as she calls it, results from

the Yeeyaners' position as members of communities of practice that often have contradicting views on the role of volunteer and professional translators in society.

In their studies, both Neather (2012) and Yu (2019) place the emphasis on the meaning that practitioners give to their experience and how that, in turn, shapes their identities. This thesis takes another stand and focuses primarily on negotiation within the context of translation standards in Wikipedia. Thus, attention is paid to negotiation as an act of mutual engagement, usually from a position of strong commitment, aimed at changing certain elements of a community's shared repertoire. Moreover, unlike Yu (2019), who seems to consider editing, proofreading and translation as sub-practices within the same community, this study situates them as contiguous practices.

As argued in 2.1.1, each Wikipedia community comprises a wide range of practices, from editing and translating to administrative tasks. In this section, the emphasis is placed on how these practices are organised within each language version and to what extent Wikipedia users engage and become full-fledged practitioners or members of their communities. In doing so, Wenger's (1998) definition of communities of practice will be used, along with concepts like 'core' versus 'peripheral' membership, taken from other practice theorists such as Shove, Pantzar, and Watson (2012). The labels 'career', to denote progression, 'core', to indicate full membership, and 'periphery', to refer to inexperienced Wikipedia users or novices, are aimed at providing a picture of how this user-driven encyclopaedia is structured from a praxeological approach. The last part of this subsection takes the question of constellations of communities one step further. It applies it to Wikipedia as a whole, as an amalgamation of multiple language versions with loosely connected practices and a set of common goals.

At the time of writing, Wikipedia has more than 97 million registered user accounts and 56 million articles or entries spread across the 323 language versions of the encyclopaedia. Of these 97 million users, only 293,423 (0.30%) were regarded as active ('Wikipedia' 2021). As explained in 2.1.1, to be considered an active user in the project, it is required to have edited a page or space at least once during the last month. This policy applies to all language versions, and in some cases – depending on the language community – it is the minimum requirement to vote for Wikipedia

administrators and other trusted roles, as well as to partake in essential decision-making processes such as introducing changes to specific policies or guidelines.

Wikipedia assesses participation and engagement by counting the number of 'edits' during a delimited timeframe of 30 days, always taking the present as the reference or point of departure. This assessment means that users with many contributions or edits to the project will be considered inactive if they have not participated recently. Nevertheless, they may still be allowed to vote or participate in decision-making processes if they have made at least 100 edits since their account was registered ('Wikipedia' 2021). To measure the degree of engagement, Wikipedia has developed a series of tools that retrieve statistical data from all its communities.

While engagement could be measured, other praxeological concepts such as membership are more ambiguous, especially in the context of Wikipedia. In principle, to be considered a member of the user-driven encyclopaedia, it is sufficient to have registered an account at some point. This broad interpretation of membership, however, only establishes a link between the user and the encyclopaedia that does not necessarily entail participation. Therefore, users could be regarded as members by virtue of being registered in Wikipedia but still be considered alien (peripheral) to their local language community or communities if they have failed to engage with others and incorporate the project policies and guidelines into their practice.

From a narrow perspective, an individual cannot be regarded as a member of a given community of practice if there is no (mutual) engagement or commitment (Wenger 1998, 75; Shove, Pantzar, and Watson 2012). The online nature of Wikipedia, however, poses some challenges to the relationship between membership and commitment, since anonymous contributors could be highly engaged during weeks or months without the benefits and obligations of membership. For this reason, and following Wenger's (1998, 74) posit that 'engagement is what defines belonging', it seems necessary to make a distinction between long-term engagement in one or more practices as a requisite for full membership, and temporary engagement as a feature shared by peripheral new editors (novices) and anonymous contributors (non-members) alike.

As a community of practice, each language version of Wikipedia has tutorial programmes and documented standards to help 'outsiders' or novices become acquainted with the project policies and progress in their careers. A few language communities even have mentorship programmes, where senior editors may volunteer to assist less wiki-conversant users ('Wikipedia:Mentorship' 2020). The encyclopaedia also has forum-like spaces known as 'talk pages' where all editors can open discussion threads, ask for assistance and, in some cases, receive supervision from experienced or senior editors. More importantly, talk pages offer a platform for editors to negotiate changes to Wikipedia articles and documents such as policies or guidelines. As will be discussed in more detail in 3.2.2, researchers in the field of multilingualism have turned their attention to the pivotal role of talk pages not only as spaces where significant decisions are reached (Hautasaari and Ishida 2012), but also as sites where potential disagreements lead to the development of hostile environments (Jones 2017; 2018a; 2018b).

To guarantee a successful inbound trajectory to full membership, users in Wikipedia are also expected to incorporate a series of documented standards to their practice. Such standards usually come in three forms: policies, guidelines and essays. According to the English Wikipedia page on the topic, policies are widely accepted and editors 'should normally follow [them]' ('Wikipedia:Policies and Guidelines' 2021). Second only to policies, guidelines are also supported by consensus and, as a result, editors are 'encouraged to follow [them]'. The third type of standard, the essay, constitutes the 'advice of an editor or group of editors' and, consequently, is not necessarily representative of the views expressed by the community. In essence, whether documented standards are classified as policies, guidelines or essays depends on the number of practitioners endorsing them.

Some documented standards, such as the policies requiring editors to write in a neutral style (WP: NPOV), and substantiate information with reliable sources (WP: VER), are found across all Wikipedia language communities. In contrast, other policies, guidelines and essays are negotiated within specific communities and they respond to the particular needs or concerns of their members. As stated in 1.1.1, differences in policy-making across Wikipedia communities can have an impact on the deployment of content-creation devices such as bots and CX. In light of the existing overlap – shared core policies – in a reduced, yet significant, number of

standards across several language versions of Wikipedia, the multilingual platform can be said to fall under Wenger's (1998, 127) definition of a 'constellation of communities'.

Figure 2-1 below shows an organisational chart in which Wikipedia, as a constellation of communities, appears on the left. This constellation breaks down in approximately 323 communities of practice, each matching a different language version of the online encyclopaedia. The chart illustrates how a typical community of practice is organised. The X label stands for any language version of Wikipedia. As explained earlier in this section, each community comprises volunteers with distinct degrees of engagement. Thus, there are registered users, new editors and senior editors. The latter are full-fledged practitioners or members of their community, while the first two groups are part of the periphery (Wenger 1998; Shove, Pantzar, and Watson 2012). Full members include administrators, bureaucrats and other functionaries, as well as experienced editors who devote little to no time to maintenance tasks. Passive users, as non-engaged members of their community, occupy a position in between the new editors (peripheral members) and the anonymous contributors.

Anonymous or non-registered contributors participate in their communities, they can modify articles and engage in discussions, but they are not regarded as members. Despite their lack of membership, their contributions cannot be underestimated. Some anonymous contributors can be and often are engaged over some time, create articles and follow the project policies. Finally, Figure 2-1, while illustrative of the hierarchical structure of Wikipedia, should be approached with caution. Unlike other communities of practice where roles are more rigid, Wikipedia contributors are volunteers whose interests and responsibilities may change over time (O'Sullivan 2009). Even editors who hold trusted – and consequently restricted – positions such as administrator (sysop) or bureaucrat might choose to devote a considerable amount of their time to other tasks such as editing articles.

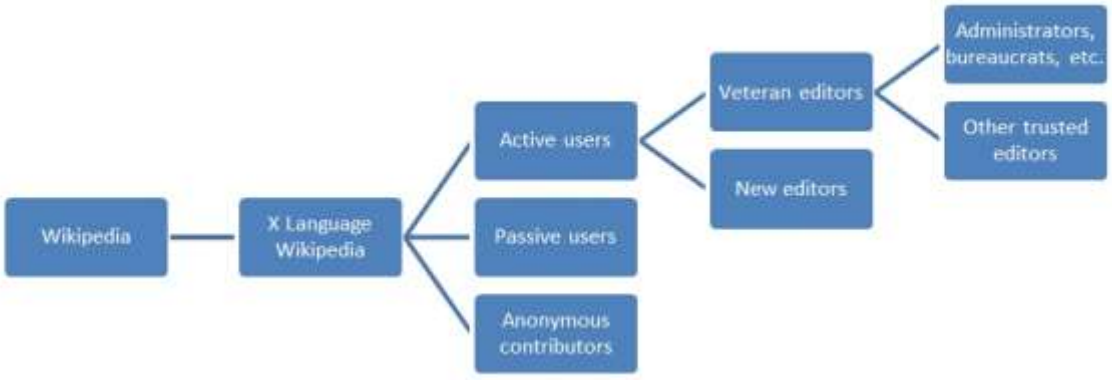


Figure 2-1. Diagram showing the typical hierarchical structure of a community of practice in Wikipedia.

## 2.3 The normative aspect of practice: the standards

As argued in the previous section, multilingual Wikipedia can be defined as a constellation of various communities of practice. Most Wikipedia language communities share a set of practices, goals and beliefs that bring them together. Nonetheless, there are standards that are unique to each community and that both regulate and give advice on how certain practices should be performed. In what follows, the normative aspect of practices is examined in more detail, paying attention to how the aforementioned concepts of engagement and membership contribute to, and rely on, the sustainment of both explicit (documented) and implicit (non-documented) rules that differentiate one community from another (Farnsworth, Kleanthous, and Wenger-Trayner 2016; Wenger 1998). In the last part of this section, the emphasis is laid on the interplay between standards and devices, preparing the ground for the study of materiality in practices, which is covered in 2.4.

Engagement in practice means that individuals agree on performing activities in a certain way, following a series of steps established by convention (Wenger 1998; Shove, Pantzar, and Watson 2012). This common set of rules, along with know-how and competence, contribute to what Warde (2016, 39) defines as 'mutual intelligibility'. He contends that all practices involve mutual intelligibility, which

is the capability to understand what is relevant to the performance of a given practice, often by observing other practitioners in action.

The amalgamation of standards, knowledge and competence set practices apart and makes it easier for both practitioners and outsiders to position themselves in relation to a specific practice. For instance, one can easily distinguish between the practices of sailing and teaching by merely looking at the individuals performing those tasks. Warde (2016, 40) enhances the role of performance in the identification of practices, for it is through 'performance that individuals carry a practice forward, expressing, affirming, reproducing and transforming it'.

According to Warde (2016, 42–43), most practices are 'irreparably normative in character' and, as such, they must have acceptable 'standards of performance'. Standards are normative in that they determine how a practice should be performed; their enforcement is ultimately the responsibility of all practitioners. Warde (2016, 42) recognises that in some cases, there are also agents tasked with ensuring compliance with the standards. In Wikipedia, for example, administrators can block users who infringe core policies or whose attitude is disruptive.

Consequently, a standard is not motivation but, instead, a threshold or parameter of effective performance. In other words, those who fail to perform practices adequately can be judged by other individuals around them. At the same time, an outstanding performance can be rewarded in one way or another, depending on the practice (Warde 2016, 45). Schatzki (2018) links normativity to accountability, which means that engaged individuals must make a concerted effort to calibrate their actions, avoiding what others may consider unacceptable, incorrect or inappropriate. In this context, proficient and compliant performers are usually positively regarded by their peers and become trusted members, an idea that is compatible with Wenger's (1998) concept of achieving full membership and Shove, Pantzar, and Watson's (2012) idea of career progression.

Standards of performance can be either implicit or explicit. For example, some rules are not written down, yet they are implicitly understood and articulated by practitioners (Warde 2016, 43). In other cases, standards have been documented and can easily be accessed by practitioners. Warde (2016, 46) argues that practices improve and evolve because throughout history there have been 'attempts at

describing and recording, for public circulation, accounts of how to do something, how to do it better, and how to do it well'.

Warde (2016) posits that these recordings are found in multiple formats, from rulebooks and instruction manuals to guidebooks and codes of conduct. Organisations, projects or institutions usually dictate formalised and codified standards, which are essential for the performance of practices. Compliance with such standards tends to be actively promoted from above. Many sports, especially at a professional level, and institutionalised professions like teaching, nursery and medicine constitute good examples of formally organised practices. As the next subsection will discuss, Wikipedia communities have documented standards of performance in the form of policies, guidelines and essays.

In conclusion, the adoption of specific standards and materials set communities of practice apart (Wenger 1998, 188). As previously mentioned, being a practitioner means that one has to follow a series of negotiated rules, whether implicit or explicit, which dictate how the practice should be performed (Wenger 1998; Warde 2016; Hui 2017). By performing certain practices and not others, and by using specific materials for the realisation of those practices, members or practitioners may identify themselves as being part of a community. This sense of belonging is built upon a set of shared beliefs and values among practitioners (Warde 2016, 46). These points of convergence, however, are not at an individual level, for human beings are construed as carriers or holders of practices (Shove, Pantzar, and Watson 2012; Warde 2016; Olohan 2017). Specifically, it is the sustained endorsement and performance of specific practices that bring people together into specific communities (Wenger 1998; Farnsworth, Kleanthous, and Wenger-Trayner 2016). With this idea as a backdrop, the next subsections provide an overview of both overarching and local Wikipedia policies and guidelines.

### 2.3.1 Standards of practice in Wikipedia: policies, guidelines and essays

> The Right Thing takes many forms, but perhaps most central is the preservation of our shared vision for the neutral point of view policy and a culture of thoughtful, diplomatic honesty.
>
> *Statement of Principles* (Wales 2021)

Wikipedia has standards that apply to all 323 language communities and which are collectively known as the 'Five pillars'. The first pillar states that Wikipedia is an encyclopaedia and not 'a soapbox, an advertising platform, a vanity press, an experiment in anarchy or democracy, an indiscriminate collection of information, or a web directory' ('Wikipedia:Five Pillars' 2021). The second explains that Wikipedia is and must be written 'from a neutral point of view', avoiding advocacy. The third and fourth make clear that Wikipedia's content is free and that everyone should treat each other with respect and civility.

The fifth and last pillar tells potential editors that the encyclopaedia 'has no firm rules' but, instead, policies and guidelines aimed at achieving some degree of objectivity. The same pillar recognises that 'the content and interpretation [of Wikipedia's guidelines] can evolve'. This variation in practices is acknowledged by most theorists (Nicolini 2012; Shove, Pantzar, and Watson 2012; Warde 2016; Hui 2017), with Hui (2017, 55) claiming that 'variations emerge through performance'. Wikipedia's fifth pillar places emphasis on learning through performance and, more specifically, on the significance of making exceptions when needed to improve the project.

Editors in Wikipedia are expected to comply with three core policies or guidelines. These policies, which fall into the overarching second pillar, are 'Neutral Point of View' (NPOV), 'Verifiability' (VER), and 'No Original Research' (NOR). The first of these documents encompasses the other two. It is defined by Wikipedia as a non-negotiable policy that 'cannot be superseded by other policies or guidelines, nor by editor consensus' ('Wikipedia:Neutral Point of View' 2021). NPOV provides examples of what should be avoided, like stating opinions as facts and advises on how to handle neutrality disputes.

VER highlights the importance of using reliable, independent sources such as university textbooks, journals, magazines and mainstream newspapers. Wikipedia also discourages original research (NOR), defined as 'facts, allegations and ideas for which no reliable, published sources exist' ('Wikipedia:No Original Research' 2021). NOR also refers briefly to other policies, most notably to those tackling plagiarism and copyright violations. The two policies provision against cases of non-attributed content and copyrighted material, warning non-compliant editors of the consequences should they decide to ignore the standards.

The policies mentioned above are universal and they apply to all Wikipedia language communities. Nonetheless, as mentioned earlier, not all documented standards in Wikipedia are ubiquitous. For instance, the Catalan Wikipedia does not include birth and death years as categories in biographical articles, whereas this is common in most other language communities ('Viquipèdia:Categorització' 2021). The Cebuano, Dutch, Swedish and Waray-Waray Wikipedia communities, among others, either lack policies or have flexible ones allowing the use of bots for the creation of articles. These variations apply to local editors, who on top of being familiar with Wikipedia's 'Five pillars' and the three aforementioned core policies, are also expected to observe the standards established by their language communities.

As will become clear in 4.1 and 4.2, translation guidelines and essays stand as elucidative examples of local documented standards. For instance, the French Wikipedia translation guidelines encourage the mobilisation of CX among editors, whereas access to the same device in the English Wikipedia is disabled for novices ('Wikipedia:Content Translation Tool' 2021). The Spanish Wikipedia has an official policy that allows experienced editors to run bots for content creation purposes under certain circumstances and subject to scrutiny before approval ('Wikipedia:Creación de artículos con bot' 2020). The Dutch and Swedish Wikipedia communities, more permissive regarding bot-creations, lack strict policies regulating this device. However, the absence of documented standards does not always result in a lax environment. An illustrative case is found in the French Wikipedia, where, despite the apparent paucity of regulation, the number of bot-generated articles remains negligible ('Wikipedia Statistics' 2019).

As the next section will show, the implementation of novel bespoke materials usually prompts changes in practices and the standards that underpin them. Moreover, other materials, which already exist and are associated with concrete standards of usage, can also acquire new scripted uses (Shove, Pantzar, and Watson 2012). The idea of materials triggering change – configuring practices – is developed further below and will be analysed in Chapter 6.

## 2.4 Materiality in and of practice

Having ascertained the importance of conventions or standards for the performance of practices, attention is now turned to the role of materials in configuring practices. Throughout this chapter, it has been argued that standards, whether documented or implicit, dictate how a practice is best performed. These standards often involve the use of materials, which may or not have a 'scripted use' or established role assigned by convention (Hui 2017). For instance, whereas objects such as hammers and pens have a scripted use (Shove, Pantzar, and Watson 2012, 47), modern mobile phones and computers offer a wide range of options, none of which are scripted.

Mobile phones can be used to call other people, take photos, share content in social networks, send text and voice messages, play games, make payments, transfer information, get directions, and so forth. As objects, they can partake in the performance of different practices, all of which have their specific conventions. Making payments with a mobile phone requires endorsing privacy agreements and following a strict protocol to guarantee a safe transaction. Conversely, obtaining directions or calling a friend are not subject to the same regulations, competences and expectations. Thus, mobile phones have taken on roles that were previously associated with other materials, such as photo cameras, credit cards, and desktop computers. Coutard and Shove (2019, 11), in their analysis of infrastructures, tie in these novel material arrangements with the demands of particular practices.

Regardless of their scripted use, materials play a significant role in the execution of practices. As the familiar tennis example illustrates, one cannot play this sport without a racket, a ball and a net. Likewise, the practice of driving to work every day would not be possible without an automobile (Shove, Pantzar, and Watson 2012). This materiality behind the performance of specific practices remains vastly

unchallenged, to the extent that most practice theorists seem to agree not only on its significance but also on its potential as a driving force of change (Wenger 1998; Nicolini 2012; Shove, Pantzar, and Watson 2012). From this perspective, Shove, Pantzar, and Watson (2012, 45) maintain that the availability of materials determines the emergence of various practices. They argue that practices consist of three separate layers: competence, access to materials, and meaning. To elucidate their point, they mention the practice of having toast for breakfast, which would be hampered, if not impossible, without a cooking appliance and bread (material), the skills to use the equipment (competence), and the idea or convention of having toast for breakfast.

In more recent work, Shove (2017) differentiates between three types of materials, depending on the function they each have in enabling the performance of a given practice. The first category is infrastructure, which Shove (2017, 156) defines as 'things in the background' that are required for the practice but do not actively engage with it. The second group consists of resources, which are the materials that are utilised and undergo changes during the execution of the practice. The third class of materials is formed by devices or 'things in action' (Shove 2017, 159; Olohan 2021, 46), which are mobilised – rather than simply used – and are directly involved in configuring the practice.

Despite these distinct functions and the ostensible pivotal role of devices as configuring elements of practice, Shove (2017, 160) posits that the three types of materials are interdependent. In brief, the performance of a practice hinges on the existence of infrastructural arrangements, resources and devices. Olohan (2021), who in earlier work highlighted the significance of 'material agency' in translation (Olohan 2014, 18), illustrates this interdependence with her example of the practice of translators. Although she acknowledges that its performance may vary depending on the materials involved, Olohan (2021, 48) holds that nowadays translators need to have access to a workplace (infrastructure), energy supply (resources), and tools (devices) for the realisation of their practice. Along the same lines, Littau (2016, 87; 2017, 97) notes that material objects such as books, computers and new technologies have played an essential role in configuring translation practices throughout history.

Other practice theorists such as Wenger (1998), who do not discriminate between the tangible – the materials – and the intangible – the concepts and styles, still recognise the significance of materiality. He situates materials as part of the shared repertoire, which, alongside mutual engagement, constitute the joint enterprise that drives each community of practice. Having a joint enterprise means that practitioners agree to do things together. This engagement would not be possible without a shared repertoire, consisting of a set of artefacts, tools, styles, and concepts that all practitioners have at their disposal (Wenger 1998, 73). Moreover, for the author, the ideas and the materials that underpin the existence of communities of practice are grouped as resources, whereas commitment – what makes individuals engage in a given practice – is at the other end of the tangible-intangible spectrum (Wenger 1998, 72–73).

Reckwitz (2002) distinguishes between bodily and mental activities on the one hand and objects on the other. For Reckwitz (2002, 252), 'objects are necessary components of many practices – just as indispensable as bodily and mental activities'. Schatzki (2010, 129) differentiates between artefacts, humans, organisms, and 'things of nature', but he includes them all under the umbrella term of 'material arrangements', which ultimately enable the enactment of practices. According to Schatzki (2010, 129), 'human coexistence inherently transpires as part of nexuses of practices and material arrangements'. Nicolini (2012), in the same line of thought as Reckwitz (2002), places emphasis on the socio-material aspect of practices. According to Nicolini (2012, 171), it is important to take into account 'the central role of artefacts and the entanglement between human and non-human performativity'. The significance of materials for the performance of practices is, in turn, reinforced by the increasing influence of technology, with mobile phones, computers and other devices changing how people communicate and engage in different practices (Nicolini 2012, 171).

As mentioned earlier in this section, the performance of practices also entails meaning. In their explanation of how new practices emerge and old ones disappear, Shove, Pantzar, and Watson (2012, 55) place meaning at the forefront of their approach, arguing that 'meanings are extended and eroded following dynamic processes of association'. Specifically, new practices are born because of the interplay between various elements whose meaning has mutated through a process

of dissociation. To elucidate this point, Shove, Pantzar, and Watson (2012, 53) refer to the practice of Nordic Walking in Finland. They posit that this sport would never have been created were not for the fact that the material components, the sticks, acquired a different meaning through a process of dissociation. In other words, sticks not only ceased to be considered tools used exclusively by individuals with reduced mobility, but they also became associated with the concept of fitness. Chapters 4 and 6 will tackle dissociation when examining the Wikipedia translators' shifting attitudes to automated devices.

The Internet has opened the doors to new markets, where individuals can buy products from other countries without leaving their households (Nicolini 2012, 171). Recent updates in automation have also influenced how translations are performed, redefining old practices. The circulation of materials, to use Shove, Pantzar, and Watson's (2012, 108) term, can be physical – transportation – or virtual. The latter can be seen in Wikipedia, where bots and devices such as CX circulate widely across several language communities. Schmidt (2017, 143) recognises that changes in materials and how they circulate determine the fate of practices. Nevertheless, he also recognises that 'changes [in practices] derive from innovations in all core elements of the bundle of practices', including cultural meanings and forms of knowledge and know-how.

Hui (2017, 57) observes that the performance of one single practice typically involves various objects, some of which transcend the practice itself. To illustrate how material components are at the crossroads of different practices, she gives the example of passports. As documents that are subject to diverse regulations, passports are issued with a set of instructions and standards that target different groups of practitioners such as government administrators, border security agents, applicants and eventual passport holders (Hui 2017, 62). In Wikipedia, one clear example is found again in the deployment of automated devices such as bots, which perform a variety of tasks that go from the most rudimentary tasks to the creation of multilingual content. This idea is developed further in 6.3.1, where the analysis targets the multi-faceted aspect of materials, which can adopt diverse roles depending on the practice they configure.

As noted at the beginning of this section, theorists recognise the relevance of materials for the performance of practices. However, discrepancy arises when it comes to their classification. For example, a few theorists such as Latour (1991), Nicolini (2012), and Shove (2017) argue that they are part of the practices themselves, while others such as Schatzki (2003) hold that they are merely external mediators that enable the performance of certain practices. At one end of the scale, those who contend that materials participate in practices on an equal footing to human beings often claim that practices are social networks in which both inter-subjective relationships among humans and heterogeneous interactions between humans and non-human actors take place (Latour 1991; Nicolini 2012). At the other end, theorists such as Schatzki (2003, 183) believe that what sets humans and materials apart is the notion of intelligibility. In this context, intelligibility means that only human actions are filled with intentionality and affectivity, both of which are necessary for the negotiation of meaning between practitioners (Nicolini 2012, 169).

It is beyond the scope of this project to discuss the theoretical implications of either of the two approaches introduced above. Notwithstanding, as was articulated in 1.2 and throughout this chapter, the relationship between practitioners and materials is one that informs the investigation, where special attention is paid to the role of materiality in the performance and configuration of translation practices in Wikipedia. The reliance on materials for the successful performance of practices in the encyclopaedia shows that, at least, in this case, the practice of translating would not be possible without access to essential resources such as power and electricity, infrastructure (Wikipedia itself) and devices. This constraint is better explained when one adheres to Shove, Pantzar, and Watson's (2012) and Shove's (2017) treatment of materials as configuring practices as opposed to external mediators (Schatzki 2003). To put it more concretely, the performance of editing and translation practices in Wikipedia not only requires access to specific materials, but their evolution depends primarily on (and is the outcome of) constant software upgrades.

### 2.4.1 Materials in Wikipedia: automated devices and Wikidata

Throughout this chapter, a point has been made that practices are imbued with materiality and that materials play an essential role in how certain practices change over time (Shove, Pantzar, and Watson 2012; Shove 2017; Olohan 2021). For these changes to transpire, however, it is necessary that different communities of practice make decisions on how to best utilise tools (Wenger 1998). Thus, considering that one of the aims of this thesis is to ascertain the impact of automation on the evolution of the translator-editors' practices in user-driven encyclopaedia, the focus is placed on two of the internal devices that have been more frequently mobilised to create multilingual content: CX and bots. Along with these devices, this subsection introduces Wikidata, a large and fast-growing database (infrastructure) launched in 2012 that, as will be argued in 6.3.1, has also contributed to configuring the practice of translation in Wikipedia.

The Content Translation Tool (CX), which was launched in 2014, is an automated device that currently serves over 100 of the 323 language communities of Wikipedia. Of these, the Spanish and French communities have long been its most active users. At the time of writing, the Spanish and French Wikipedias have approximately 92,000 and 74,000 CX-generated articles, respectively ('Content Translation Statistics' 2021). Furthermore, statistical data retrieved from the tool's page show that English is by far the primary source language of the translations. Not only do the Spanish and French communities stand out in the number of articles created using CX, but they also occupy the first and second position regarding the number of editors that have mobilised the device since it was first implemented seven years ago. As of August 2021, 13,991 Spanish Wikipedia editors and 10,580 French Wikipedia editors have used CX at least once ('Content Translation Statistics' 2021).

As noted by Laxström, Giner, and Thottingal (2015), and McDonough Dolmaya (2017), CX was conceived with the specific purpose of encouraging and facilitating translations across different communities of Wikipedia. The success of CX probably resides in the fact that it is a user-friendly interface for translators that may not be familiar with wikicode. To begin with, CX users are required to choose the Wikipedia article they want to translate. Thereafter, a separate window pops up,

showing the ST on the left side of the screen. The right side, which is blank at the outset, is for users to work on the target text (TT). The TT is automatically generated by the device as users click on the blank half of the screen. Similar to what happens with other CAT tools, the TT mirrors the structure of the ST paragraph by paragraph. Once the text has been generated, translators can modify terms or expressions from the TT if, in their view, they have not been rendered correctly. The device also allows translators to save unfinished articles in the platform for up to two years without uploading them to Wikipedia until they think they are ready for publication.

The software is updated periodically, with more improvements being introduced to facilitate the translating experience. One distinctive feature of this device is that it uses algorithms to suggest potential Wikipedia articles for translation. Thus, these algorithms retrieve information from the editor's latest translations through Wikidata and recommend a series of articles on the same topic. Recent updates to the device also warn translators against potential translation shortcomings and give advice on how to address them. Figure 2-2 below illustrates how an article is translated using CX.

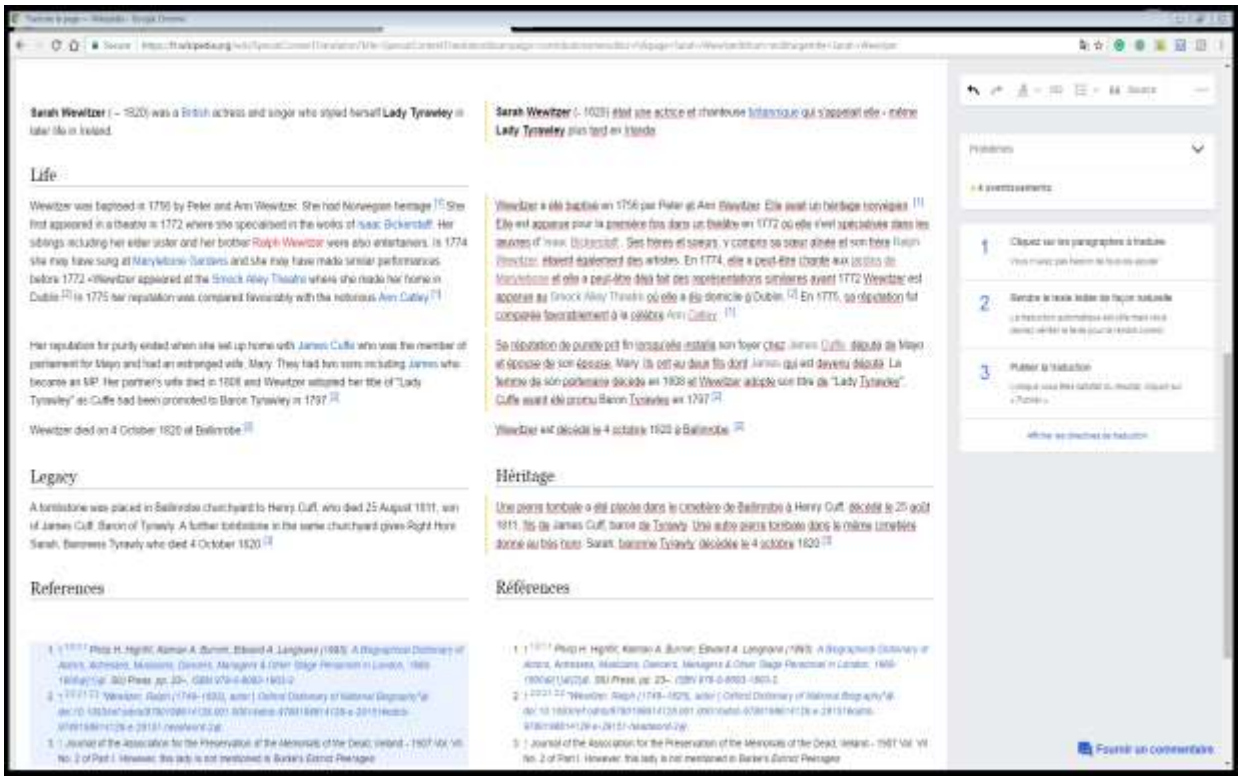

Figure 2-2. Screenshot showing the ongoing translation into French of the English Wikipedia article on the British actress Sarah Wewitzer (-1820) using CX.

As stated in 1.1.2, a software robot or 'bot', unlike the content translator, is an automated device designed to perform a wide range of tasks in Wikipedia. Bots are usually deployed to detect instances of plagiarism, orthographic errors, and illegitimate edits or 'vandalism' to Wikipedia pages made by disruptive users (Lih 2009; Geiger 2011; O'Hagan 2016; Tsvetkova et al. 2017). Moreover, as explained in the previous chapter, besides these repetitive tasks, bots can, in some cases, be used to create short Wikipedia articles or entries.

At the time of writing, not all 323 language communities of Wikipedia allow bots to be used for this purpose. Those that have taken a more lenient approach are often communities with a small number of registered editors such as the Asturian, Cebuano, Chechen, Dutch, Serbian, Swedish, Urdu, Volapuk and Waray-Waray Wikipedia projects. Wikipedia communities with larger numbers of editors tend to have stricter bot policies regulating and restricting the mobilisation of these devices. As will be explained in more detail in 4.2.3 and 4.3, in these latter communities owning a bot is restricted to experienced Wikipedia editors or full-fledged practitioners, subject to the scrutiny and approval of the community. Since bots can

produce mass creations and have the potential of causing more disruption if they malfunction, bot runners must be versed in coding and enjoy the trust of their Wikipedia community.

Unlike CX, bots do not ‘translate’ from one language into another. Instead, they operate with a series of commands and templates programmed by their owners. These templates consist of pieces of information and quantitative data that are introduced into the program, a process known as parsing, which enables the bot to create fairly standardised Wikipedia articles or entries, typically about subjects dealing with figures or statistics. Figure 2-3 provides an example of the quintessential bot-generated article in the Swedish Wikipedia.

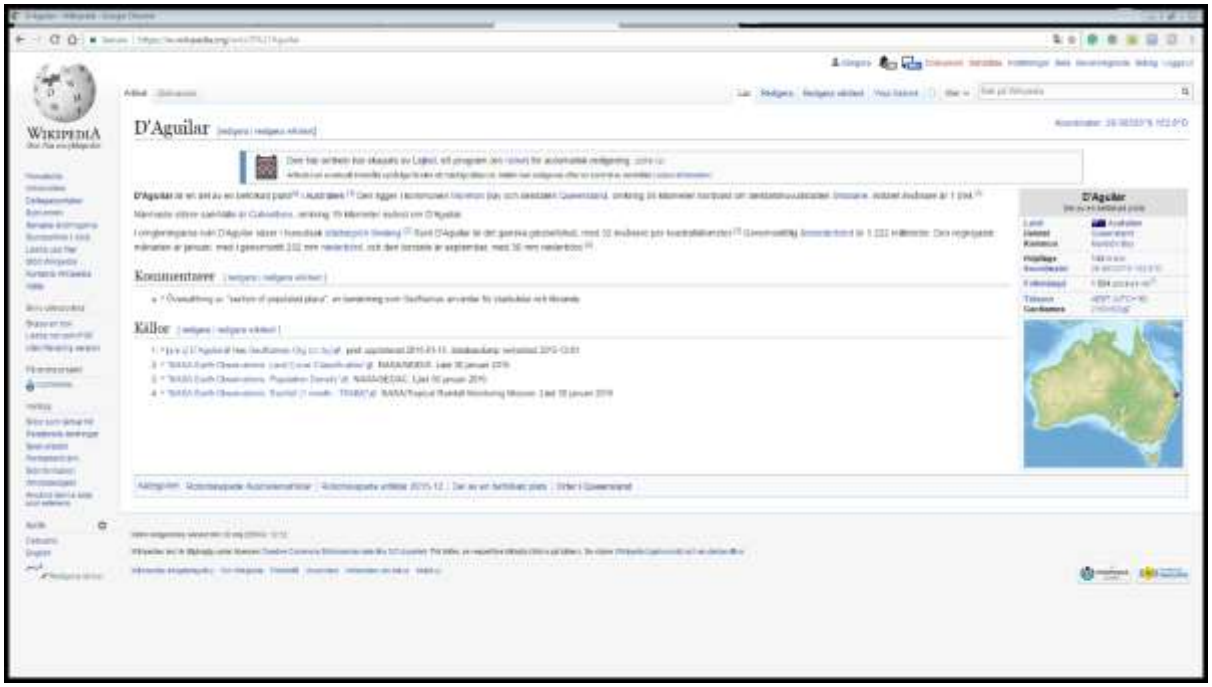


Figure 2-3. Screenshot of the Swedish Wikipedia bot-generated article on the Australian town D’Aguilar.

Having introduced the two primary devices that will be examined in this thesis, attention is now turned to Wikidata. As its name suggests, Wikidata is a multilingual collaborative repository. First launched in 2012, the repository is ‘document-oriented’, and it comprises almost 95 million instances of structured data known as ‘items’, each identified with a unique number or code (‘Wikidata Main Page’ 2021). All items are prefixed with the letter Q, to identify ‘the topic the item covers’ and represent it ‘without favouring any language’. Similar to what one

would expect to find in an archive, a typical Wikidata item consists of a label, a description, and a series of statements.

For the sake of illustration, the item about the University of Manchester ('Q230899' 2021) contains the label 'University of Manchester', the description 'public research university in Manchester, England', 18 statements (instance of 'university', logo image, photo, inception, country, coordinate location, member of X, affiliation, subsidiary, owner of X, etc.), and 27 identifiers (GeoNames ID, VIAF ID, Art UK Venue ID, Quora topic ID, Twitter username, etc.). Statements are records containing all the information available about an item, while identifiers are lists of websites and other external sources with which the item is or has been associated.

As is the case with Wikipedia articles, items on Wikidata are also interconnected and categorised. The item Q230899 on the University of Manchester has a wide range of embedded items, from 'university' (Q3918) and 'Manchester' (Q21525592), to 'Universities Research Association' (Q4005852) and 'United Kingdom' (Q145). Wikidata items serve all the language communities of Wikipedia as well as other sister projects such as Wikimedia Commons, an online repository of free-use audio-visual material. Thus, most Wikipedia articles, regardless of the language in which they are written, are linked to a unique Wikidata item. For example, the Wikidata item about the University of Manchester currently serves 59 language communities of Wikipedia. This means that at least part of the information contained in the Wikidata item is displayed on the information box (infobox)[8] of all the Wikipedia language communities that have an article about the university. Figure 2-4 below shows the Wikidata item Q230899 with some of its statements.

[8] Infoboxes or infobox templates are panels, usually located at the right of a Wikipedia article, either at the top in the desktop version or at the bottom in the mobile app. They provide a summary of the subject matter covered in the article. Besides text, infoboxes may contain other multimodal content such as images and maps. See Figure 2-3.

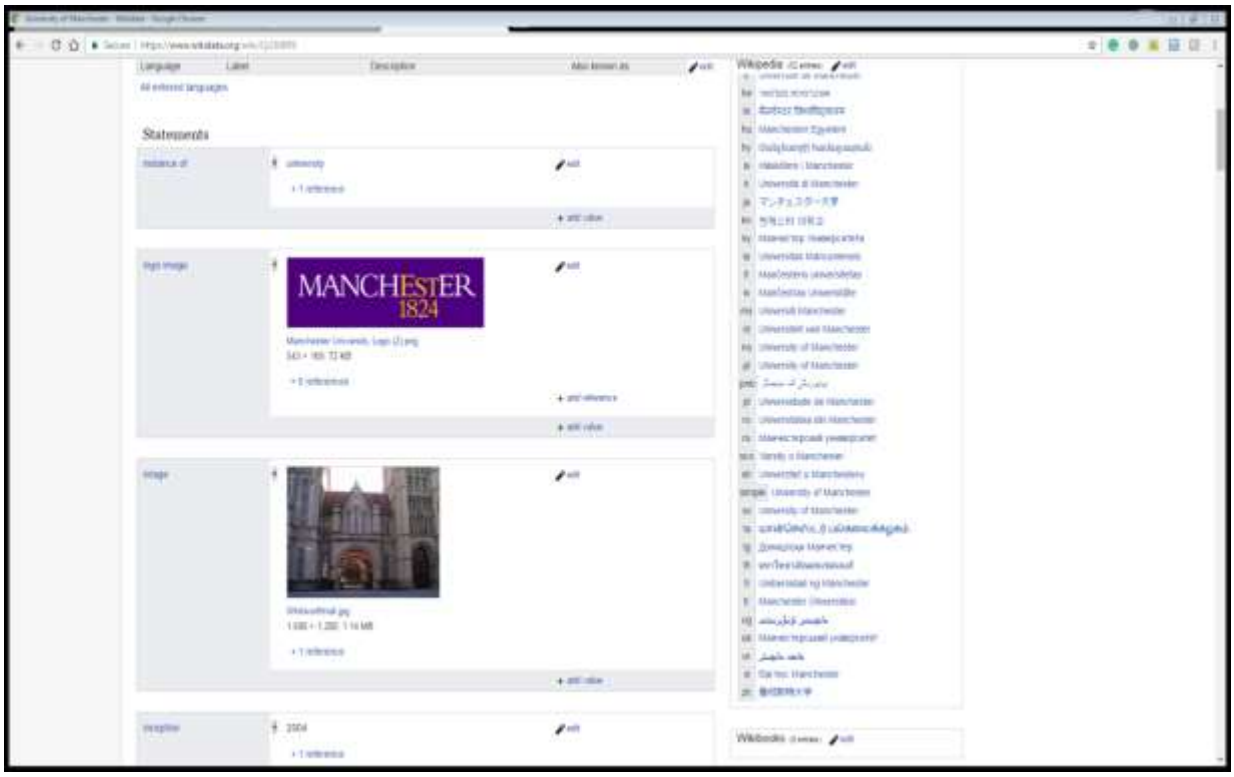


Figure 2-4. Wikidata item on the University of Manchester showing some of its statements.

## 2.5 Chapter conclusion

This chapter introduced the theoretical principles that sustain the investigation. Drawing on practice-theoretical concepts, the emphasis was placed on four key aspects: a) what is understood by practice (Wenger 1998; Reckwitz 2002; Nicolini 2012), b) what constitutes a community of practice (Wenger 1998; Wenger, McDermott, and Snyder 2002), c) how most practices have a series of standards of performance (Warde 2016), and d) how materials not only enable the realisation of practices but also configure them (Shove, Pantzar, and Watson 2012; Shove 2017).

All four sections of the chapter set the ground for the study of translation practices in Wikipedia. Section 2.1 provided an overview of some Wikipedia practices that are shared by most language communities. Section 2.2, which tackled communities of practice, framed each individual language community of Wikipedia as a distinct community of practice and argued that Wenger's (1998) concepts and terminology can be applied to their study. In so doing, this section provided the conceptual tools

that are necessary to describe how Wikipedia editors of the four language communities negotiate, challenge and endorse certain policies and guidelines.

To better understand how standards operate in Wikipedia, 2.3 discussed a few Wikipedia policies, guidelines, and essays. This section highlighted the relevance of documented standards in Wikipedia, not only to the study of practices but also to gain some perspective on how translation practices are negotiated within individual language communities. Finally, 2.4 built on the concepts introduced in the previous sections and centred on the essential role of materials in configuring practices.

The next chapter presents the methods that will be used to collect and analyse the data.

# Chapter 3. Methods and data selection

This chapter introduces the methodological framework that was constructed for the study. Section 3.1 outlines the three primary analytical stages that inform the analysis of translation practices and how these are intertwined. Section 3.2 goes on to describe the first of these methods: the analysis of a set of documented standards in the form of Wikipedia policies, guidelines and essays. Section 3.3 presents the second phase, which consists of a thematic analysis of comments posted primarily on the ancillary talk pages of the documents examined in the first phase. Then, section 3.4 tackles the use of semi-structured interviews as an effective method to gather information on, among other things, the extent to which 16 experienced Wikipedia translators have assimilated the standards and how the mobilisation of content-creation devices has configured their practice. All sections establish relevant links, when appropriate, to practice theory and previous research on Wikipedia. Sections 3.2 and 3.3 are divided into three parts: a) introduction, b) data collection and analysis, and c) methodological challenges and ethical considerations. Section 3.4 follows a similar structure but includes another subsection devoted to explaining aspects of the research design. Finally, the last part of the chapter contains a summary of the main points covered, thereby preparing the ground for the investigation of translation practices in chapters 4 to 6.

## 3.1 Analytical stages

The first of the analytical stages comprises the thematic analysis of documented standards on translation and content-creation devices (policies, guidelines and essays) in each of the four Wikipedia language communities under investigation. Since the first aim of this study is to understand how translation practices have been regulated in individual language communities of the encyclopaedia, it is therefore essential to start the research with a fine-grained analysis of the standards that seem to underpin their performance. As discussed in 2.2 and 2.3, standards are usually negotiated, agreed and ultimately enforced by members of a given community of practice (Wenger 1998). Consequently, in this case their study is a good way to gain insight into the arrangements that different Wikipedia communities have put in place to guide translators. This examination of documented standards is then

followed by an analysis of translation-related comments posted by engaged editors on the ancillary forum-like platforms known as 'talk pages'. In doing so, the objective of the first two analytical stages is to investigate the extent to which translation practices are regulated (RQ1) and standards negotiated (RQ2) in the four Wikipedia language communities.

Following the analysis of documented standards and their negotiation, a series of one-to-one semi-structured interviews are conducted to understand whether and, if so, how experienced Wikipedia translators of the four communities have incorporated the standards into their practice (RQ3). Moreover, as stated in 1.2 and 1.3, another core aim of this study is to ascertain the impact of automation and metadata in configuring the translators' practices over the past five years (RQ4). Therefore, participants were asked about their use of and views on automated devices such as CX and bots. The sections below provide further information about the three analytical stages as well as how the data were collected.

## 3.2 Analysing normativity: the documented standards

As mentioned in 2.3.1, Wikipedia comprises a series of policies and guidelines that are diverse in nature and scope. Some policies, such as those featuring in the 'Five pillars', are common to, and consequently enforced by, all the 323 Wikipedia language communities. Others, dealing with translation and the use of automation, particularly bots, tend to be more localised and thus vary from one community to another. This distinction between the universality and locality of policies and guidelines has enabled the classification of Wikipedia as a constellation of several communities of practice, each matching a different language version of the encyclopaedia (2.2). Moreover, as became apparent in 2.2.1, documented standards in Wikipedia follow a tripartite framework. These three components, in decreasing order of communal consensus and enforceability, are policies, guidelines (sometimes referred to as 'help pages') and essays.

Another main point asserted in Chapter 2 is that most practices are intrinsically normative (Wenger 1998; Warde 2016). As such, practices have standards that determine how individuals should perform them. When it comes to Wikipedia, both universal and local policies and guidelines set a benchmark for good practice.

Considering that policies, and to some extent guidelines, are approved by consensus, editors are expected to be compliant, to the extent that those who fail to do so can be banned from the platform in some cases, depending on the seriousness of the offence (Morgan and Zachry 2010, 168). On a more general level, however, policies and guidelines are devised to facilitate the editing process, prompt users to engage in the project, and ensure civility and respect among Wikipedia contributors.

Over the last decade, an increasing number of scholars working in other domains of the humanities and social sciences have turned their attention to the study of Wikipedia institutional regulations. Morgan and Zachry (2010), for instance, have conducted research on the role of policies, guidelines and essays in negotiating changes to content in Wikipedia. They analysed a dataset consisting of 47 policies, 232 guidelines, and 404 essays from the English Wikipedia with special focus on those that were often cited on talk pages. Their findings revealed that English Wikipedia editors usually cite policies and guidelines in their interactions with other peers, thus suggesting that policies in the encyclopaedia have a significant regulatory role (Morgan and Zachry 2010, 168).

Along the same lines, Bilić's (2015) study of social construction of knowledge in Wikipedia shed light on the significance of policies and guidelines in informing decision-making processes. According to Bilić (2015, 1258), consensus in Wikipedia is reached through 'the routinization of the process [of mediated content production] in policies and guidelines'. In line with this argument, a more recent investigation on collaborative processes among Wikipedia editors with polarized political views has shown that policies and guidelines are invoked by both sides to challenge each other's views (Shi et al. 2019, 332).

Against this backdrop, as explained in 1.2, this thesis incorporates the analysis of policies, guidelines and essays to the study of translation practices across the Spanish, French, Dutch and Swedish language communities of the user-driven encyclopaedia. In particular, the first part of the investigation draws primarily upon a thematic analysis of all existing documented standards of translation and automation collected from the four communities. The documents from the aforementioned communities were first scrutinised in search of themes related to different aspects of practice. Those themes that appeared in two or more Wikipedia

communities were classified as ‘universal’ (overlap) for the purposes of the investigation, whereas those that featured in one community were considered ‘local’.

As explained in chapters 1 and 2, the study of documented standards can help understand how distinct communities of practice have attempted to regulate translation and the deployment of content-creation devices such as CX and bots. While it is true that everyone translates differently, even within the same community, the existence of local policies, guidelines and essays is intended to set the minimum standards of practice for a particular language version of the encyclopaedia (‘Wikipedia:Policies and Guidelines’ 2021). Therefore, examining explicit standards is a first step to understand how individual Wikipedia communities view and approach translation.

### 3.2.1 Data collection and analysis

The extant data[9] selected for the analysis of documented standards come from a small subset of Wikipedia policies, guidelines and essays of the four communities under investigation. These documents, which differ from Wikipedia articles in their layout and content, are not part of the main space of the encyclopaedia. Instead, they are highly institutionalised technical spaces that dictate, or give advice on, how to best perform a practice. As part of their inbound trajectory (Wenger 1998), editors are expected to familiarise themselves with the precepts of most core policies, while they should ‘attempt to follow’ the guidelines (‘Wikipedia:Policies and Guidelines’ 2021). To find these pages, editors can browse them using a search engine or through ‘Category: Wikipedia policies and guidelines’. Figure 3-1 below, taken from Wikipedia’s overarching ‘Five pillars’ for illustrative purposes, shows two categories: a) Wikipedia basic information and b) Wikipedia policies and guidelines. By clicking on each one of them, users are redirected to a space that hosts a myriad of similar pages under the same category.[10]

[9] As Salmons (2016) notes, ‘extant data’ are data generated without the direct influence of the researcher. This type of data usually includes documents, postings and other online material.
[10] Another option is to click on the Wikipedia policies and guidelines template at the bottom of the page, also shown in Figure 3-1.


Figure 3-1. Screenshot of the Wikipedia policy 'Five pillars'.

Considering that this thesis aims to study translation alongside the deployment of content-creation devices in the context of multilingual Wikipedia, it is therefore necessary to first analyse and acknowledge the normative – and by extension, explicit – aspects that may underpin the editors' practices and regulate access to automation. Consequently, the dataset comprises a subset of all existing policies, guidelines and essays on translation and content-creation devices (CX and bots), as available in the Spanish, French, Dutch, and Swedish language communities of the online encyclopaedia at the time of writing.[11] Such a selection excludes cross-wiki editing policies, which admittedly play a significant role in the creation of encyclopaedic articles. This exclusion is however justified by the fact that this study has set out to investigate translation as opposed to other forms of editing, even though previous studies have argued that both form a continuum in Wikipedia (Jones 2017; 2018b). To put it more concretely, whilst translation in the encyclopaedia can be considered a subtype of editing, not all editing involves translation. Accordingly, the analysis in 4.1 and 4.2 targets a specific subset of documents that address translation and the mobilisation of content-creation devices.

[11] As shown in Table 3-1 below, the Swedish Wikipedia does not have translation policies and guidelines. Instead, the community has an essay on translation.

Table 3-1 below presents the documented standards selected for analysis. Of the 13 documents, ten have the status of guidelines (G), two have been classified as policies (P), and one is an essay (E). As the table illustrates, the Swedish Wikipedia occupies a unique position in that it lacks policies and guidelines on translation. At the time of conducting the analysis, the community had only one essay entitled *Översättningsrekommendationer* [Translation recommendations] giving advice on how to perform the practice. Although less adherence to its recommendations is somewhat expected given the lower status of essays (see 2.2.1), the page was included in the sample because it was the only documented standard available. This situation, however, changed during the course of the study when a second essay was posted. On 16th July 2021, following a short-lived discussion on the ancillary talk page, the original essay was renamed *Wikipedia:Rekommendationer vid översättning från engelska* [Recommendations for translating from English] ('Wikipedia:Rekommendationer' 2021). A second essay, now bearing the name of the previous essay, was drafted shortly thereafter to give more general advice on how to translate from languages other than English. Thus, the current version of *Översättningsrekommendationer*, which is not examined in this thesis, is based on the document first published on 22nd January 2021 under the name of *Översättning* [Translation].

The table also shows the date when each document was published in Wikipedia for the first time, as well as the latest version that was selected for the analysis.[12] Thus, modifications or post-edits occurring after June 2020 do not form part of the dataset. Wikipedia's volatility as a drawback in data analysis has been reported in previous studies (Jones 2017; 2018b). This limitation, coupled with the fact that the examination of the standards lays the foundation for the semi-structured interviews, made it necessary to concentrate exclusively on the latest version of each document at the time of commencing the analysis.

[12] Both the date of publication and the latest version of each document were retrieved from the revision history.

| Community | Page title and status | Title in English | First published | Latest version |
|---|---|---|---|---|
| Spanish | *Wikipedia:Traducciones* (G) | Wikipedia:Translations | 28/04/2005 | 23/11/2017 |
| | *Ayuda:Cómo traducir un artículo* (G) | Help: How to translate an article | 08/09/2008 | 24/06/2020 |
| | *Ayuda:Traducción de contenidos* (G) | Help: Content translation | 24/08/2016 | 09/04/2020 |
| | *Wikipedia: Creación de artículos con bot* (P) | Wikipedia:Creating articles with bots | 01/04/2007 | 06/08/2014 |
| French | *Aide:Traduction* (G) | Help:Translation | 05/11/2007 | 24/05/2020 |
| | *Wikipédia:Traduction automatique* (G) | Wikipedia:Automatic translation | 26/02/2010 | 21/10/2018 |
| | *Aide:Outil de traduction* (G) | Help:Translation tool | 31/03/2015 | 14/04/2020 |
| | *Aide:Pywikipedia* (G) | Help:Pywikipedia | 26/03/2006 | 22/04/2020 |
| Dutch | *Help: Tips voor het vertalen van een artikel vanaf een andere Wikipedia* (G) | Help:Tips for translating an article from another Wikipedia | 10/10/2006 | 21/08/2019 |
| | *Wikipedia:Content Translation* (G) | Wikipedia:Content Translation | 29/05/2015 | 25/01/2018 |
| | *Help: Gebruik van bots* (G) | Help:Using bots | 20/12/2004 | 28/04/2020 |
| Swedish | (†) *Wikipedia:Översättnings Rekommendationer* (E) | Wikipedia:Translation recommendations/advice | 04/01/2007 | 08/04/2020 |
| | *Wikipedia:Robotar* (P) | Wikipedia:Bots | 31/05/2005 | 10/02/2019 |

Table 3-1. Documented standards selected for the study grouped by Wikipedia language community.

All the documented standards were translated into English and saved in separate files to facilitate their analysis. For documents written in languages other than Spanish, the original plan was to work alongside a group of Wikipedia translators. This decision was taken to minimise the risk of misinterpreting the textual data due to a lack of proficiency in Dutch, French and Swedish. Moreover, given the fact that a few documents exhibit a high degree of technicality – commonly referred to as the 'Wikipedia jargon' – which may pose challenges to a layperson, having volunteers conversant with the rules of the site was considered an advantage. This

approach proved successful with French and Swedish Wikipedia contributors. For the Dutch Wikipedia, however, the lack of available volunteers at the time prompted the search of an alternative solution. Thus, in this latter case the translation was altruistically performed by a fellow researcher who was also a native speaker of the language.

Once translated into English, the data collected from the Wikipedia pages were probed and coded, placing emphasis on the specificities of local policies, guidelines and essays. In particular, the thematic analysis in 4.1 and 4.2 draws upon data on aspects of translation practice that were subsequently coded following criteria such as 'overlap' and 'dissonance'. These criteria surfaced after a process of coding and theme identification that was conducted inductively. As noted by Saldaña (2021, 41), inductive coding occurs when the list of themes emerges from examining the data. Thus, this process is opposite to deductive coding, where the researcher starts their project with a set list of themes (Saldaña 2021, 40).

The first part of the analysis is based on themes that surfaced across various Wikipedia communities (universality). For instance, the requirement to verify translated content with reliable sources (WP:VER) was mentioned across documented standards of the four communities and thus constitutes an overarching theme. The second phase of the analysis (4.2) examines themes that featured in one or two communities (locality). By way of illustration, the documents give slightly different advice on how to choose an article for translation (4.2.1). While the Spanish and French Wikipedia guidelines prioritise the quality of the ST, their Dutch and Swedish counterparts place the emphasis on being familiar with the text one wants to translate. As a result, two separate themes were identified: a) Quality and b) Familiarity.

### 3.2.2 Methodological challenges

The thematic analysis of textual data presented minor, albeit significant, challenges. For example, identifying a theme was not always a straightforward process. As Bryman (2016, 587) and King, Horrocks, and Brooks (2019, 200) note, there is no agreement among scholars as to what constitutes a theme. Nevertheless, most seem to agree on the fact that the repetition of a concept in a text is not sufficient to have

a theme if such concept proves irrelevant to the investigation. Thus, one fundamental aspect that was considered in processing the data was the suitability of certain codes to provide meaningful information (theme) in response to the issues raised by RQ1. An initial screening soon revealed the gross disparity between the selected documents. Some provided much detail whereas others, particularly *Help:Gebruik van bots* and *Wikipedia:Robotar*, contained information that was mostly geared towards the deployment of bots for mundane (non-content-related) maintenance tasks. A similar imbalance was observed in the Spanish page *Wikipedia:Traducciones*, which turned out to be a brief document that echoed advice found in *Ayuda:Cómo traducir un artículo*.

Another point worth noticing is that although working side by side with translators from the encyclopaedia was aimed at minimising the risk of misinterpreting the data due to a lack of fluency in a given language, that decision inevitably means that a considerable part of the analysis draws upon material translated by a third party. While this can be a problem in some cases, Wikipedia documents tend to be written in a formal register that leaves little room for alternative readings. Besides, email correspondence remained open during and after the translation process, so that all parties could ask questions or request clarification if and when they were needed.

Alongside these challenges, there were others associated with the research design. For example, Wikipedia's volatility (Jones 2017; Shuttleworth 2018) and the fact that each analytical stage builds on the previous one required the investigation to be narrowed down to a specific timeframe, with the subsequent omission of data published after the round of interviews took place between July and August 2020. All things considered, it is important to acknowledge that, despite their potential for mutability, the core aspects of the policies, guidelines and essays remain, for the most part, vastly unchanged. This relatively slow shift occurs because, as noted in 2.2 and 2.3, significant alterations to policies and guidelines tend to be the outcome of a process of negotiation and, as such, they require communal consensus. Thus, as will become clear in 4.3, to amend policies – and to some extent, guidelines – editors must come together by a process of mutual engagement and agree on the proposed changes on the talk pages. Section 3.3 below explains how the analysis of negotiation is addressed and undertaken in the third part of Chapter 4.

## 3.3 Analysing the negotiation of standards in talk pages

The second analytical stage consists of a thematic analysis of translation-related comments, when available, posted on the Wikipedia talk pages attached to the documented standards in the previous section. By examining how editors of the four communities challenge and negotiate their local standards over time, this dataset contributes to answering RQ2. In essence, the analysis of comments or postings in discussion threads targets the negotiability of practices directly by drawing attention to instances of agreement and dissent (rebellion) around translation and the mobilisation of content-creation devices. In so doing, it paves the way for the study of the translators' incorporation of the standards and the impact of automation in configuring their practices, both of which are covered in chapters 5 and 6 respectively.

'Talk pages', as their name suggests, are forum-like spaces attached to a Wikipedia article or document in which contributors can comment on matters concerning that specific page. Anyone can access the ancillary talk page by clicking on the tab 'Talk' (or its equivalent depending on the language community) located near the top left corner of the page in question. As one of the main communication channels between Wikipedia contributors, talk pages play an essential role in the process of volunteer co-production and community building. Consequently, their study can help understand, for instance, how editors engage with one another and handle contentious issues (Jones 2017; Góngora-Goloubintseff 2020). For further elucidation, Figure 3-2 below shows a screenshot of the talk page attached to the English Wikipedia document 'Translate us' ('Wikipedia Talk:Translate Us' 2020).

## Translating medical content [edit]

I am in discussions with Translators Without Borders [1]@. Both they and I are interested in working on a collaboration to translate medical articles first into simple English and then into other languages. Are there others here interested in helping? Please visit here Wikipedia_talk:WikiProject_Medicine#Medical_translation if you are. **Doc James** (talk · contribs · email) 03:22, 15 December 2011 (UTC)

## Expand templates [edit]

It would be useful to have a list of the equivalents of the {{Expand}} that exist on other language wikipedia sites. While I may not be able to translate the English version to that language, or even necessarily read it, it is pretty obvious when there is a relative stub or red link in the other language. Being able to place the that language's equivalent of {{Expand English}} on the page would be useful.

--J Clear (talk) 13:35, 7 July 2013 (UTC)

## Templates for acknowledging copying from interwiki pages [edit]

I wondered if here would be an appropriate place to ask a question like this. From my experience in editing on the English Wikipedia, I have come across a template that looks like this: "This article may be expanded with text translated from the corresponding article in the German/Hungarian/French/Spanish (etc) Wikipedia." When you add information to an English stub article, for example, from another language, you add the template to the Talk page that specifies what language you have copied content from. I wonder if the same applies to editors of other languages?

To help make this clearer I'll explain what exactly I am asking: I have recently had article Rocks and Honey nominated and passed as a GA. A few months before, the Hungarian version was passed as a FA. While overseeing the review process, one reviewer recommended that the editor of the article follow a similar style of writing as on the English Wikipedia page, by covering critical reviews of the album both positive and negative (at the time the Hungarian version only had positive reviews mentioned). The editor then copied the English content word for word using a machine translator (Google Translate). I am aware that when articles contain information translated from other languages a template from Wikipedia:RFT should be added to the article's Talk page. Does this also apply on other languages? Even if so, I don't believe that there is a version of that template in Hungarian. Bonnietylersave (talk) 16:23, 25 July 2013 (UTC)

## It does not say where to do the actual translation [edit]

Figure 3-2. Screenshot of the English Wikipedia talk page attached to document 'Translate us'.

Talk pages have drawn the attention of researchers from various fields (Schneider, Passant, and Breslin 2010; Laniado et al. 2011; Ferschke, Gurevych, and Chebotar 2012; Gredel 2017). In studies on multilingualism, Hautasaari and Ishida (2012) were amongst the first to collect data from Wikipedia talk pages. Their dataset consisted of 228 talk pages from the Finnish Wikipedia, 93 from the French Wikipedia, and 94 from the Japanese Wikipedia (Hautasaari and Ishida 2012, 59). All the talk pages included in their investigation belonged to Wikipedia entries that had been translated into the three aforementioned languages. Upon analysing comments posted by Wikipedia editors on those pages, Hautasaari and Ishida (2012, 57) discovered that most discussions in the three language communities centred on the translation of proper nouns, source referencing and transliteration. More recently, Jones' (2018a; 2018b) research on the English Wikipedia entries for Tokyo and Paris probed into content posted on the ancillary talk pages to better understand how editors proposed, endorsed or opposed changes to the narrative.

In light of the social value of talk pages, the analysis conducted in 4.3 has as its primary focus a series of translation-related comments – including the mobilisation of automated devices – posted by editors of the four Wikipedia language communities. The aim is to ascertain how the documented standards examined in 4.1 and 4.2 are negotiated (RQ2). Thus, the thematic analysis of talk pages and

similar forums within each Wikipedia language community can provide insight into the underlying processes of mutual engagement in general and instances of rebellion or dissent in particular. At their core, talk pages are social spaces where Wikipedia editors interact with fellow practitioners and debate on topics of their interest. For this reason, an examination of comments posted on these pages can reveal how aspects of the shared repertoire are, and have been, contested by individuals in the pursuit of their community's joint enterprise. The analysis of the data draws upon the theoretical tools introduced by Wenger (1998) and Wenger, McDermott, and Snyder (2002), with particular attention paid to the concepts of 'community disorders' and 'rebellion'. The findings are then triangulated in search of points of convergence and dissonance between the language communities.

### 3.3.1 Data collection and analysis

As Table 3-2 below illustrates, the data for this part of the analysis were gathered from discussion threads posted on the aforementioned talk pages prior to June 2020. In those cases where the information provided in the discussion threads was obscure or required further clarification, experienced editors who had previously been involved in the debate were approached via private correspondence using Wikipedia's 'email this user' function, which features on the navigation bar at the left of the user page.[13] The original dataset comprised 13 talk pages, one per documented standard. Nevertheless, after an initial screening, six pages were eventually discarded because the information they contained either was insufficient or was beyond the scope of the study; this is indicated in bold (Table 3-2). Following this latter criterion, three other pages were chosen to supplement information extracted from the sample: a) *Wikipedia:Meldingen botimport nieuwe artikelen* [Wikipedia:Notifying of new bot-imported articles], b) *Användare:Lsj* [User:Lsj], and c) *Användarediskussion:Lsj* [User talk:Lsj]. Although not part of the main dataset, these pages were consulted during the analysis to substantiate certain claims, particularly those regarding the mobilisation of bots.

[13] Only registered Wikipedia editors who enabled the 'email this user' function will display it on the navigation bar.

Concerning the selected seven talk pages, the focus was placed on all translation-related comments posted since the documents were first published in Wikipedia. This decision was not only motivated by the fact that the number of discussion threads published from 2015 onwards is scant, but also because the inclusion of earlier comments helped to contextualise the current scenario. For instance, a major turning point in the history of the Dutch Wikipedia occurred in 2005 when the community gave green light to the creation of articles by bots. Similarly, in 2008, the Spanish Wikipedia editors approved a restrictive use of bots for exactly the same purposes. Thus, ignoring the underlying motivations and the issues raised by these editors would have led to a contextual gap in the investigation.

| Community | Page title | Title in English | First published | Latest version |
|---|---|---|---|---|
| Spanish | *Wikipedia discusión: Traducciones* | Wikipedia talk:Translations | 27/12/2005 | 03/03/2009 |
| | ***Ayuda discusión:Cómo traducir un artículo*** | Help talk: How to translate an article | 10/06/2010 | 17/01/2018 |
| | *Ayuda discusión:Traducción de contenidos* | Help talk: Content translation | 28/03/2019 | 08/04/2020 |
| | ***Wikipedia discusión: Creación de artículos con bot*** | Wikipedia talk:Creating articles with bots | 02/04/2007 | 30/08/2015 |
| French | ***Discussion aide:Traduction*** | Help talk:Translation | 07/05/2015 | 17/05/2020 |
| | ***Discussion Wikipédia:Traduction automatique*** | Wikipedia talk:Automatic translation | 26/02/2010 | 12/02/2020 |
| | ***Discussion aide:Outil de traduction*** | Help talk:Translation tool | 17/04//2015 | 17/03/2020 |
| | *Discussion aide:Pywikipedia* | Help talk:Pywikipedia | 17/04/2007 | 19/01/2011 |
| Dutch | ***Overleg help: Tips voor het vertalen van een artikel vanaf een andere Wikipedia*** | Help talk:Tips for translating an article from another Wikipedia | 02/09/2007 | 03/01/2015 |
| | *Overleg Wikipedia:Content Translation* | Wikipedia talk:Content Translation | 29/05/2015 | 10/06/2015 |
| | *Overleg help: Gebruik van bots* | Help talk:Using bots | 05/02/2005 | 20/02/2018 |
| Swedish | ***Wikipediadiskussion:Översättnings rekommendationer*** | Wikipedia talk:Translation recommendations | 17/05/2007 | 16/11/2018 |
| | *Wikipediadiskussion:Robotar* | Wikipedia talk:Bots | 20/05/2005 | 14/11/2016 |

Table 3-2. List of talk pages selected for analysis.

In preparation for the analysis, the comments were exported to four separate files, one per Wikipedia language community. The texts were rendered into English by the same individuals who performed the translations of the standards for the previous analytical stage. To ease the workload of the volunteers, the discussion threads sent out for translation were structurally and deductively pre-coded taking into account the issues raised by RQ2.[14] The editors who participated in the discussions were generally senior registered members of their language communities. The registration date of each editor in Wikipedia was retrieved from the page statistics tab, accessible through the revision history of their user page. Furthermore, as will become clear in 4.3, the date of publication of individual comments or threads was gathered from the timestamps attached to the editors' signature.[15]

Since the negotiation of standards on talk pages transpires within the confines of individual Wikipedia communities, the analysis in 4.3 presents the findings per language version. It does so by following the ISO[16] order of categorisation. Thus, 4.3.1 concentrates on Spanish (ES), 4.3.2 on French (FR), 4.3.3 on Dutch (NL), and 4.3.4 on Swedish (SV). The editors' usernames were also anonymised following the same criterion, which is 'ISO code X', where X stands for the order in which each participant is cited in the analysis, e.g. ES 1 [Spanish Wikipedia editor 1], NL 2 [Dutch Wikipedia editor 2]. Despite there being no ethical barriers against using the editors' usernames, even more so considering that the comments on talk pages are publicly accessible, a decision was made to semi-anonymise the data. This approach differs from the one employed in previous studies (Jones 2017; 2018a), where editors were often referred to by their usernames in Wikipedia. The semi-anonymisation of the data was considered a more suitable option in this study because the practitioners' identities are far less relevant (Shove, Pantzar, and Watson 2012; Olohan 2017) than their seniority and capacity to engage with other

[14] Having identified the main thematic areas in the first analytical phase, this second stage focused on the changes proposed to those critical areas. A structural approach to coding (Saldaña 2021, 130) was adopted to further investigate the commonalities and differences revealed by the data gathered from the talk pages.

[15] Anyone who posts comments on Wikipedia is required to sign their contribution using the 'signature and timestamp' function available on the Visual Editor. Alternatively, editors working on wikicode have to sign their messages with four tildes (~~~~), which will then show both their user name and the timestamp once their contribution (edit) has been posted.

[16] ISO stands for 'International Organisation for Standardisation'.

members of the community. Consequently, the editors cited throughout 4.3 are first presented with their year of registration in Wikipedia, while their comments are accompanied with a timestamp for the sake of transparency.

Finally, to ascertain whether the documented standards examined in 4.1 and 4.2 underwent changes as a result of negotiation, the analysis also drew upon 'revision histories'. These tools, sometimes referred to as 'edit histories', are records containing detailed information on all the changes ('revisions') undergone by a Wikipedia page over time. These changes can be traced chronologically by clicking on the 'View history' tab located at the top of each Wikipedia page. As Figure 3-3 below illustrates, every change is marked with a timestamp, the author's user name (or IP address if not registered), and the size of the modification indicated by the number of bytes (McDonough Dolmaya 2015; Jones 2017). In addition, each revision may contain an 'edit summary' where Wikipedia editors briefly describe the modifications they introduced to the page.

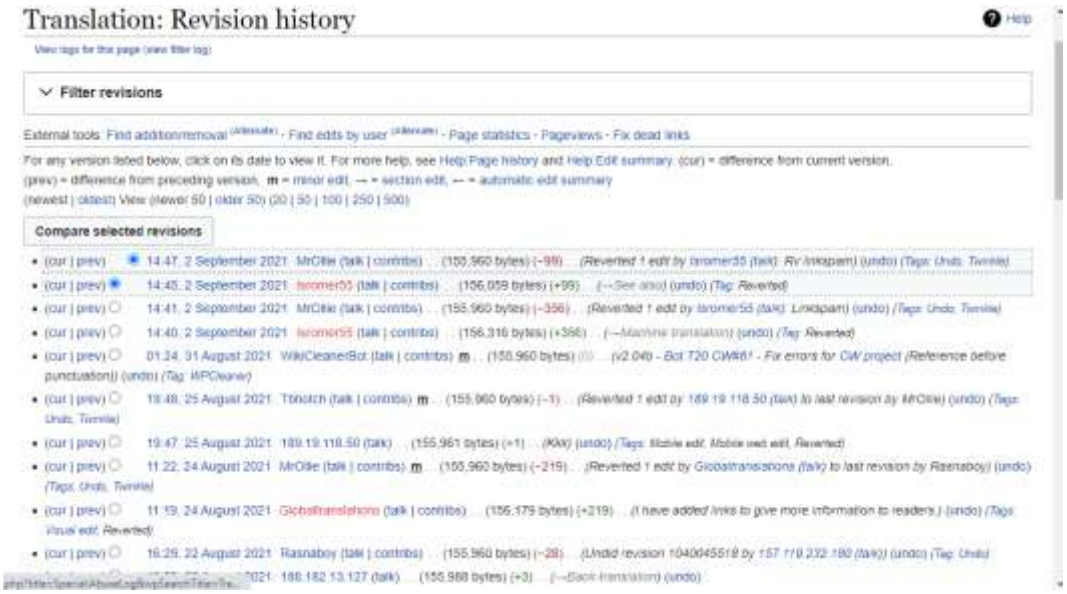


Figure 3-3. Revision history of the English Wikipedia article 'Translation'.

### 3.3.2 Methodological challenges and ethical considerations

The analysis of extant data on talk pages presented similar challenges to the ones already described in 3.2.2. First, as was the case with the examination of the documented standards, the interpretation of the textual data gathered from the discussion threads relied on translated content. Unlike the standards, however, the

rendering of online posts into English posed additional obstacles to the volunteer translators. This is because, as noted by Gredel (2017), postings tend to be written in a language that is more relaxed in style and thus display conversation-like features that are context-dependent. Moreover, because the interpretation of comments vastly depends on the interplay of pragmatic and semantic elements, it was not always possible to understand what all editors meant in concrete situations. This being the case, special care was taken when analysing the comments. In cases where the situation required it, editors were approached privately to clarify messages with an unclear context or antecedent. In other cases, as stated in 3.3.1, ambiguous threads were excluded from the sample.

Compliant with the current ethical regulations established by the University of Manchester, the examination of comments on talk pages is not subject to the same constraints that affect research with human participants. However, some early and now dated approaches to 'netnography' – the observation and study of online phenomena – had a tendency to be more restrictive concerning the selection and handling of this type of data. Eysenbach and Till (2001, 1103), for example, observed that 'internet community members do not expect to be research subjects', and even though consent is normally not needed, researchers should endeavour to analyse the data without causing grievance. Even more stringent, Kozinets (2002, 65) argued that research involving online postings should not be undertaken unless consent has been granted by the authors whose comments were selected for analysis.

Langer and Beckman (2005) have dismissed Kozinets' (2002) proposal, adding that, if anything, such restrictions should only apply to private online communities. Along the same lines, Roberts (2015, 316) contends that the study of online postings is an effective way of circumventing the procedural ethics process required for interviews, although, similar to Eysenbach and Till (2001), she acknowledges the importance of being aware of potential ethical issues arising from the selection of the data, especially when investigating vulnerable groups. Cognizant of these implications, the study conducted in this thesis draws upon public data gathered from postings whereby editors discuss topics that lie exclusively within the remit of their role as engaged members of Wikipedia.

One last point to consider, also stemming from the sample, concerns generalisability or size, a topic which will be covered in more detail in 3.4. Admittedly, the data used for this part of the analysis represent the viewpoints upheld by a minority of editors with a keen interest in translation and the mobilisation of content-creation devices. Although this may come across as a small number, even more so when Wikipedia's size hinders generalisations based solely on selected samples of qualitative data, the postings that form the basis for the analysis provide sufficient information to address the issues raised by RQ2. As Saldanha and O'Brien (2013, 36) observe, non-generalisable findings can still 'make contributions to knowledge beyond the particular' if they are able to explore questions of *how* and *why*. This being the case, the data used for the analysis of negotiation in Wikipedia do not intend to provide a definitive all-encompassing answer to this question; instead, they aim to shed light on the extent to which a specific subset of documented standards are negotiated across the four language communities under investigation.

## 3.4 Interviews

The third and last dataset used for the study of translation practices in Wikipedia comes from one-to-one semi-structured online interviews with 16 experienced editors of the user-driven encyclopaedia. The criteria adopted for the selection of participants, including what counts as experience, are discussed in more detail in 3.4.1. For now, it is worth noting that the interviews were expected to provide an invaluable insight into those aspects of translation practices that cannot be covered by a thematic analysis of documents and postings. For instance, issues such as how individual editors handle content-creation devices can only be addressed by undertaking one-to-one interviews with Wikipedia volunteers who have been actively involved in translation activities in their respective language communities.

As one of the most frequently used methods in qualitative research (Mann 2016), interviews are suitable for eliciting information about someone's background and experience. Thus, while from a praxeological lens the individuals' identities are far less relevant than their capacity to perform their practice and engage with their peers, the first part of the interviews targets the participants' experience. In particular, Chapter 5 begins by discussing the participants' motivation(s) to translate (5.2) and

the challenges they had to overcome (5.3) throughout their inbound trajectory (Wenger 1998) to set the ground for a holistic analysis of their incorporation of the standards (5.4-5.5) as well as their views on and expertise in mobilising content-creation devices, which are examined thereafter in Chapter 6.

As became apparent in 3.2 and 3.3, the analysis of documents and postings or discussion threads provides an effective way to understand the extent to which translation practices have been regulated and negotiated in specific Wikipedia language communities. Nevertheless, an additional layer of analysis is necessary to ascertain whether experienced translators have incorporated the standards established by their communities into their practices (RQ3), and to assess how their usage of automated devices and metadata has contributed to the evolution of translation in Wikipedia in the last lustrum (RQ4). In keeping with these aims, the interviews were designed to: a) gather information on the participants' assimilation of as well as disregard for particular standards; b) help determine whether translation practices, as performed by the participants, rely on both explicitly formulated documents and implicit courses of action (expectations); c) gain a better understanding of how the participants' deployment of content-creation devices (materials) and metadata have configured translation practices in Wikipedia; and d) shed light on the future of translation in the encyclopaedia by asking the participants about the changes they envisage.

Interviews have been used with relative frequency by practice theorists (Wenger 1998; Nicolini 2012; Olohan 2021). Their usage has enabled researchers to gain access to specific communities of practice, often to complement real-time observations of practices in a wide range of settings such as healthcare centres (Nicolini 2011; Maller 2015), company offices (Yli-Kauhaluoma and Pantzar 2016), schools (Lynch et al. 2017), and translation agencies (Olohan 2021). Against this backdrop, the interviews in this thesis seek to gain deeper understanding of the standards that underpin the practices of volunteer translators of four language communities of Wikipedia. In essence, the interviews can help to pinpoint issues in ways that would not be possible by simply observing individuals performing their practice. The subsection below provides further information about the data selection criteria and the format adopted in the design and realisation of the interviews.

### 3.4.1 Data collection and analysis

The interviews were conducted primarily in English between July and August 2020. Participants were selected based on the following criteria: a) demonstrable experience working as Wikipedia translators, b) fluency in English or Spanish, and c) access to Zoom, Skype or similar. The number of participants (16) was deemed to be fairly representative and manageable given the limited timeframe to conduct the interviews and the disruption caused by the Covid-19 pandemic, thus increasing the likelihood of completion. Questions about the 'right' number of participants are recurrent in the qualitative research literature. Mann (2016, 79), for instance, contends that many qualitative studies tend to have between six and 12 interviewees. Kvale and Brinkmann (2009, 113) elevate the number to approximately 15, but recognise that it can vary depending on the purpose of the investigation and the resources available to researchers. Other scholars such as Bryman (2016, 416) avoid giving figures and argue that the number is determined by saturation[17], which can be difficult to anticipate.

The participants in this study were selected following purposive sampling criteria, which means that only those that met certain criteria were considered for the study. To answer the research questions and better understand how translation practices have changed in the last lustrum, participants were expected to be experienced (+1,000 edits) in Wikipedia, with a registered account since at least 2017 (three years at the time of the interviews). In addition, they were required to have translated no fewer than 10 articles within their language communities. Having the benchmark of at least 10 translations within three years was important to ensure a certain amount of experience and familiarity with Wikipedia policies, guidelines and essays. The candidates' contributions, the number of articles they translated, and their date of registration were retrieved from the revision history of their user pages.

The selection process originally involved the participation of local Wikimedia groups, also known as 'chapters', which acted as gatekeepers (Jupp 2006, 1; Bryman 2016, 78; King, Horrocks, and Brooks 2019, 58). Wikimedia chapters are

[17] Saturation occurs when new qualitative data cannot yield more information.

registered non-profit organisations that operate in a specific country or territory on behalf of Wikimedia Foundation. Chapters are a good first option for recruiting participants, not only because they are institutionalised and trusted by most Wikipedia volunteers, but also because arguably chapter members tend to know better which editors meet the selection criteria. After several attempts were made, however, only three of the 16 participants – all from the French Wikipedia – were recruited through their local chapter Wikimedia France. The remaining participants were recruited privately using the 'email this user' function available in their personal page in Wikipedia. A total number of 34 candidates were approached via this latter method, with a response rate of 32.3% (11 participants). A third communication channel, the 'village pump'[18], was therefore considered to recruit the last two participants, both from the Dutch Wikipedia, required to complete the cohort. Since it is a public platform, the village pump was exclusively used to make the announcement. Thus, the candidates that replied to the call did so using the same 'email this user' function introduced above.

The two other criteria set for the selection of participants were imposed by the layout of the interviews, the researcher's linguistic skills and the resources available. Thus, as explained at the beginning of this section, participants were required to be fluent in either spoken English or Spanish and have access to Zoom or similar software. Except for three participants, who requested to be interviewed in Spanish, the rest felt comfortable speaking English. Similarly, most participants agreed to be interviewed on Zoom. Efforts were made to accommodate the demands of two participants who objected to being interviewed on Zoom for personal reasons. In these cases, the free and open-source multiplatform voice application Jitsi was used instead.

Upon expressing their interest in the project, all participants were sent a follow-up email containing three files. The first one was a document explaining the research aims, data management procedures, and the rights and duties they had as participants, including the right to withdraw at any time should they wish to do so

[18] Wikipedia language communities have a main forum known as the 'village pump'. Village pumps are very similar to talk pages, but unlike the latter they are not attached to a particular Wikipedia page. They are spaces where members of the community can make announcements, initiate threads addressing a wide range of topics, and discuss general policies.

(see Appendix I). The second file was an informed consent sheet (see Appendix II) that participants were asked to complete and sign if they agreed to the terms laid out in the first document. The third and last document was a questionnaire designed to gather some background information such as a) level of education, b) age group (18-30, 31-40, 41-50, >50), c) level of English[19], and d) experience in translation (either as volunteers or professionals) before joining Wikipedia (see Appendix IV). In light of its briefness and specificity, the questionnaire was only aimed at collecting statistical data, ensuring that all selection requirements were fulfilled, and avoiding asking personal questions during the interview stage.

All interviews, which lasted an average of 45 to 70 minutes, were audio-recorded on both Zoom/Jitsi and a bespoke back-up device provided by the university. In compliance with ethical regulations governing research with human participants, the recordings were then encrypted using 7-Zip and saved safely using the secure institutional storage system allocated to postgraduate researchers and staff members for research data. The data underwent a process of anonymisation – participants were given fictitious gender-neutral names – and transcriptions were commissioned to an independent organisation, which was required to sign a confidentiality agreement. To further protect the privacy of the participants, the password-protected online platform ZendTo, recommended by the university, was chosen to share files between the researcher and the transcribers. This option was considered a better alternative to Dropbox and Google Drive.

NVivo12 was used for coding the data. As noted by Bryman (2016, 602) and King, Horrocks, and Brooks (2019, 204), Computer Assisted Qualitative Data Analysis Software (CAQDAS) can be useful to process large amounts of data. For instance, NVivo12 allows researchers to upload interview transcripts to the software and identify potential themes as they work on the text. One way of doing this is by selecting a specific section of the text and assigning a theme to it. The result is a list of themes, each linking to different parts of the transcripts that can be retrieved and serve as evidence for the analysis. Thus, in this study NVivo12 proved to be a valuable tool to help keep the data organised and categorised. This feature is

[19] Participants were asked to self-rate their knowledge of English using the Wikipedia Babel categorisation scheme: 1 (basic), 2 (intermediate), 3 (advanced), 4 (near native), 5 (professional), and 6 (native, usually marked as 'N' in Wikipedia).

particularly important in studies involving a sizeable number of participants, an aspect that is discussed further below.

The coding process of the data was primarily deductive and structural, although sub-themes were also identified inductively and sometimes even simultaneously. To put it more succinctly, a list of preliminary codes emerged from the main thematic areas covered in the interview guide (see 3.4.2), which were structurally devised to gather data aimed at addressing the issues raised by RQ3 and RQ4. Therefore, themes such as 'Important policies and guidelines [each participant prioritises]' and 'Automation [pros and cons]' were already established. Nevertheless, data elicited from the interviews led to the proliferation of sub-themes that were coded following an inductive approach. As explained in 3.2.1, inductive coding is a data-driven process. Thus, sub-themes such as 'Verifiability of Content (WP: VER)' and the 'Five pillars' were defined only after examining the participants' answers.

Finally, since coding occurred in different stages, where some themes had to be refined, the process can also be considered cyclical (Saldaña 2021, 88). In this process of theme identification, there were a few instances of simultaneous or co-occurrence coding, where one response could potentially cover a primary sub-theme and a range of other sub-themes (Saldaña 2021, 124). For example, some participants that discussed the advantages to deploying bots would often justify their reasoning either by citing specific policies or by reflecting on their practice as translators.

### 3.4.2 Thematic areas: interview guide and piloting

The interviews were conducted using a guide (Appendix III), which comprised 14 questions divided into four thematic blocks: a) background or experience (1-6), b) policies and guidelines (7-9), c) automation (10-11), and d) the evolution of translation in Wikipedia (12-13). Although all questions were aimed at eliciting specific information, the semi-structured nature of the interviews allowed participants to elaborate on their answers and give examples. Likewise, cues were employed on occasion to guide interviewees who either did not fully understand the question(s) being asked or were vague in their response. These cues usually came

in the form of follow-up questions and were based primarily on the participants' answers. To put it more concretely, cues were used to help participants elaborate on certain claims or statements that could be relevant to the study.

The interviews fall within the category of 'low risk', which means that formal ethics approval was not required at the University of Manchester. This exemption is due to the fact that participants were interviewed in a professional capacity, as practitioners of a specialised practice, and thus sensitive questions were avoided altogether. Besides, in compliance with the university regulations governing interviews under the low risk category, the video-call feature on Zoom/Jitsi was disabled during the audio-recording process ('University Ethical Approval' 2021).

As Table 3-3 below illustrates, the first (questions 1 to 6) of the four thematic blocks of the interview guide targets the participants' experience as translators in Wikipedia, whereas the second section (7-8) comprises questions about the platform's standards (policies, guidelines and essays). Question 9, on the use of external devices, marks a transition to the material aspect of practice, which is the focus of the third thematic area (9-11) tackling the deployment of automation (bots and CX). The fourth and last thematic block (12-13) begins by addressing the evolution of translation practices in the user-driven encyclopaedia, and ends by asking participants to share their views on the future of the practice they perform, with the emphasis on the potential impact of Wikidata. The last question (14), which is not included in any of the thematic blocks, was designed to allow participants to further elaborate on other aspects of their practice and bring to the surface any topic that, in their opinion, was worthy of attention.

| Theme or focus | Question areas | Aim | Key theoretical concepts |
|---|---|---|---|
| Experience | 1-6 | To elicit information about the participants' experience as Wikipedia editors and translators. | Socialisation into the practice: journey from newcomer/peripheral to full-fledged practitioner (Wenger 1998, Shove, Pantzar, and Watson 2012). |
| Wikipedia policies and guidelines | 7-8 | To gain an insight into how much the participants' know about the standards that underpin their practices. | Standards of performance (Warde 2005, 2016).<br>Shared repertoire, joint enterprise and rebellion (Wenger 1998).<br>Implicit and explicit elements of practice. |
| Mobilisation of automated devices | 9-11 | To obtain information on the participants' views on and usage of external and internal content-creations devices such as bots and CX. | Materiality in and of practice (Shove, Pantzar, and Watson 2012, Shove 2017, Olohan 2021). |
| Evolution of translation in Wikipedia | 12-13 | To better understand the extent to which participants are aware of significant changes in their practice over the last years either as a result of mobilising devices, changing standards or both. | Mutability of practice (Wenger 1998, Shove, Pantzar, and Watson 2012, Olohan 2021). |

Table 3-3. List of primary thematic areas of the interview, linked to key theoretical concepts.

The questions in this part of the interview were aimed at eliciting information on the participants' motivations and reasons to become volunteer translators, the challenges they had to overcome along their journey from newcomers (or peripheral members) to full-fledged practitioners (Shove, Pantzar, and Watson 2012), the criteria they apply when selecting an article for translation, and their frequent areas of interest. Although some of these questions may come across as individual-

oriented, they were devised to find common trends, ascertain how the participants were socialised into the practice, and establish superordinate themes.

In light of the fact that someone's experience as a newcomer can have an impact on the decisions they make as senior practitioners (Wenger 1998; Shove, Pantzar, and Watson 2012), the first thematic block of the interviews was deemed necessary not only to learn more about the participants, but also to provide a more robust basis for research into their incorporation of the documented standards. Thus, the second round of questions (7-8) sought to find answers to the issues raised by RQ3; in particular, a) the significance of normativity (Warde 2005; 2016) for the performance of translation practices in the online encyclopaedia, and b) the participants' position regarding their community's shared repertoire (Wenger 1998). To this end, participants were asked which, in their view, were the most critical Wikipedia policies and guidelines that translators should follow, and which, out of those mentioned, they tended to prioritise. This open-ended question allowed participants to comment on any policy or guideline that was relevant to them. Cues were sometimes used to obtain further information on these documented standards.

Having ascertained the relevance of the standards, the third block of the interview was designed to collect data on the participants' views on, and mobilisation of, content-creation devices. As explained in 2.4, materials in general, and devices in particular, play a consequential role in configuring practices (Shove, Pantzar, and Watson 2012; Shove 2017; Olohan 2021). Thus, this section attempted to explore the underlying weightiness of both external and internal devices (bots and CX) in shaping the practice in Wikipedia. Participants were encouraged to share their views on bot-generated content and explain whether, in their opinion, the device had been – or could be – beneficial or detrimental to their language community. Shortly thereafter, they were asked about their experience using automatic translation, particularly Wikipedia's bespoke CX. Since the objective of this block was to gather data to answer RQ4, special attention was paid to the advantages and downsides to deploying this device.

The fourth and last block of questions (12-13) aimed to bring together the normative and the material components of practice covered in the previous sections. Specifically, participants were asked to reflect on the most fundamental changes

they had witnessed in their Wikipedia community over the last years. By probing into the participants' accounts of what had transpired in their community, the objective was to extract additional data to better understand the evolution of practice (Shove, Pantzar, and Watson 2012; Olohan 2021). The evolutionary dimension of translation in Wikipedia was further explored in Question 12, wherein participants were encouraged to share their views on the future of their practice.

Finally, a piloting session was run in June 2020 with an experienced translator of the Asturian Wikipedia who met the sampling criteria. Piloting is a common procedure in qualitative research and is intended to test the comprehensibility and appropriateness of the questions so that any errors can be amended prior to the commencement of the official interviews (King, Horrocks, and Brooks 2019, 91). Saldanha and O'Brien (2013, 158) observe that piloting is an essential part of the interview design because it allows researchers to test aspects such as the wording and clarity of the questions. They further contend that although pilot testing is not always effective with semi-structured interviews due to their unpredictability, it can still help researchers to make significant changes to the questions, learn from trial and error, and assess their performance (Saldanha and O'Brien 2013, 178). As far as this study is concerned, piloting allowed the researcher to implement minor adjustments to a few questions. Despite having only one volunteer, the interview offered a good opportunity to gain experience and test the recording features on Zoom and the device that was used as a backup.

### 3.4.3 Methodological challenges

This last subsection discusses a series of challenges stemming from the sampling criteria, as well as issues of generalisation and transferability, and validity threats. For instance, the choice of English or Spanish as vehicular languages for the interviews inevitably raises some problems that, as insignificant as they may seem, cannot be overlooked. To a lesser extent, the requirement to have synchronous communication channels such as Zoom and a good internet connection also imposed constraints on the selection of participants. More generally, there are a series of additional challenges that are inherent to interviews, such as power balance

differences between interviewer and interviewee and ethical dilemmas that cannot be ignored like the Hawthorne effect.

The first issue, generalisation or generalisability – the potential of the dataset to be extrapolated to the wider community – was discussed, albeit briefly, in 3.2 and 3.3. Numerous scholars have referred to generalisability as one of the most common objections raised against qualitative research methods, especially interviews (Kvale and Brinkmann 2009; Saldanha and O'Brien 2013; Bryman 2016; Mann 2016). According to Kvale and Brinkmann (2009, 261), the claim that data collected from interviews are not generalisable rests on the misconception that scientific knowledge is universal. In this regard, they argue that context plays a significant role in the production and dissemination of knowledge in the humanities, and that 'a specific interview situation may [still] be transferred to other relevant situations' (Kvale and Brinkmann 2009, 262).

Closely related to context is the idea of scope. Saldanha and O'Brien (2013, 36), in their approach to semi-structured interviews, observe that failure to ponder the scope of the research findings can weaken their validity. In the case of Wikipedia, a user-driven project with thousands of contributors spread across 323 languages, generalisation through semi-structured interviewing can only be achieved within the specific scope of the investigation. As noted by Leung (2015, 325), the specificity of qualitative studies makes generalisation to larger groups unnecessary. Thus, while the findings from the interviews cannot be extrapolated to all Wikipedia translators, they can still be indicative of *how* translators of individual language communities of the encyclopaedia perform translation as well as the devices they mobilise to achieve their goals.

As anticipated at the beginning of this section, another problem derives from the sampling criteria chosen for the study. In particular, the requirement that all participants communicate in English or Spanish excluded potential candidates that may otherwise have been in a good position to answer the interview questions in their language. Even if participants are able to speak English or Spanish fluently, some researchers argue that it is desirable to let interviewees express themselves in their first language (Saldanha and O'Brien 2013; Mann 2016). Although there is a strong case for conducting the interviews in the participants' primary language, this

is not always possible for a number of reasons. As regards this study, limited competence in Dutch, French and Swedish on the researcher's side led to the choice of either English or Spanish as linguae francae. This limitation, however, did not have an impact on the duration of the interviews. Although participants who were using English as a lingua franca occasionally struggled to find the English translation of specific terms, there were no substantial differences between the length of the English interviews and those in which Spanish speakers used their language.

Another problem that was considered when preparing the interviews was the requirement for participants to have access to Zoom, Skype, Google Hangouts or similar software. As King, Horrocks, and Brooks (2019, 121) note, while platforms such as Zoom or Skype shorten the distance between researcher and remote participants, its correct functioning requires fast broadband. A bad internet connection may not only hamper communication but it can also render some parts of the recording inaudible. Another downside of using Zoom and other synchronous communication channels is that they may deter potential participants from taking part in the interviews if they are not familiar with the software (King, Horrocks, and Brooks 2019, 122).

Although the above challenges were originally taken into consideration, the outbreak of Covid-19 changed the global landscape, resulting in a surge in demand for platforms such as Zoom as more people were forced to work remotely. Given these unprecedented circumstances and the restrictions imposed on international travel, synchronous communication channels became the only viable option to undertake the interviews. In the case of this study, barring some exceptions, communication between the researcher and the interviewees was generally not hindered by the factors considered in the planning phase. Besides, not using video may have enhanced the quality of connections.

Besides the challenges already mentioned, there are ethical issues to consider. The first one concerns the reliability of the participants' responses. Some qualitative researchers have drawn attention to the so-called 'Hawthorne effect' (Kvale and Brinkmann 2009; Saldanha and O'Brien 2013). Saldanha and O'Brien (2013, 153), for example, posit that this occurs when participants intentionally alter their

behaviour and instead of expressing their sincere opinion on a certain topic, they provide the answer they think will satisfy the researcher. Although this is difficult to anticipate and control (Mann 2016; King, Horrocks, and Brooks 2019), steps were taken to avoid leading questions, and participants were actively encouraged to speak freely with the reassurance that their data would later be anonymised.

The last, albeit important, ethical issue that inevitably arises is connected to the researcher's knowledge of the subject as a long-time Wikipedia editor and translator. While some authors argue that the 'insider' status of the researcher can influence the participants' answers negatively, one major advantage is that participants are less likely to simplify ideas and can therefore provide an elaborate account of their practice (Mann 2016; King, Horrocks, and Brooks 2019). Moreover, from a praxeological perspective, being perceived as a fellow practitioner can pave the way for a more relaxed conversation wherein participants do not feel observed by an 'intruder' or 'outsider' (Wenger 1998). As noted in 2.2.1, this 'insider' approach has also been adopted in previous studies on volunteer translators as communities of practice (Yu 2019).

## 3.5 Chapter conclusion

This chapter has presented the analytical stages that will guide the investigation. In particular, it has demonstrated how the study of translation practices in Wikipedia can be undertaken from various angles, using distinct datasets to triangulate the data. Different sections of this chapter have also delineated the criteria for the selection and collection of the data and explained how certain choices inevitably engender significant methodological challenges. As a result, some space has been devoted to elucidating how these hurdles were overcome. Having introduced the methodological framework, chapters 4 to 6 move on to analyse the data.

# Chapter 4. Documented standards and their negotiation

The aim of this chapter is twofold. First, it seeks to analyse translation standards across the four language communities under investigation to determine to what extent they inform each community's practices and reveal their local expectations following internal decision-making processes. In doing so, this first stage is aimed at answering RQ1: 'How and to what extent have the four Wikipedia language communities regulated translation practices?' To this end, attention will be paid to how certain standards are shared across communities while others stem from internal decisions made by individual Wikipedia communities.

Secondly, upon analysing the standards, the next step is to examine how Wikipedia editors from the foregoing communities mutually engage in the translation and bot debates that transpire in the related talk pages. Within this second aim, the focus will be placed on how editors negotiate – endorse and challenge – specific normative aspects that dictate, or provide advice on, how translation should be best performed in their respective Wikipedia language communities. This analytical phase tackles RQ2: 'How are translation standards negotiated within each of the four communities?'

As stated in chapters 2 and 3, members of a given community of practice – also known as practitioners – come together by what Wenger (1998) defines as a process of mutual engagement, characterised by the existence of a joint enterprise. This enterprise, in turn, is supported by a set of beliefs, standards, concepts and materials known as the shared repertoire or common domain (Wenger 1998; Wenger, McDermott, and Snyder 2002; Farnsworth, Kleanthous, and Wenger-Trayner 2016). The prevalence of an underlying common goal or interest is therefore crucial for the performance of practices. This driving force has been defended even by those practice theorists (Nicolini 2012; Shove, Pantzar, and Watson 2012; Buch and Schatzki 2019) who do not subscribe to Wenger's (1998) concept of community.

In what follows, the normative aspect of Wikipedia's translation practices is brought to the forefront. The peer-produced, user-driven encyclopaedia is a constellation of hundreds of language communities, each of which is largely

characterised by having a set of specific documented standards resulting from internal processes of mutual engagement and negotiation. Against this background, the next sections analyse both overarching and specific standards that underpin the performance of translation practices and how these are negotiated, agreed and contested by engaged practitioners on talk pages.

## 4.1 Overlap across policies: Wikipedia as a constellation

In 2.2.1, Wikipedia was defined as a constellation of communities of practice. Core to the concept of constellation is the idea that certain communities of practice are loosely connected by common values, a shared history, and related enterprises (Wenger 1998, 127; Wenger, McDermott, and Snyder 2002). In other words, for constellations to exist there must be an element of continuity across two or more communities of practice. According to Wenger (1998, 127), this continuity is present in discourses and practices that transcend single communities. In the case of Wikipedia, each of the 323 language communities has common goals and stands for the same values, i.e. the distribution of free, neutral, and reliable knowledge.

One elucidative example of these shared values appears in the definition and scope of Wikipedia. The four language communities under investigation refer to Wikipedia as 'the free encyclopaedia' on their main page. In the case of the French, Spanish and Swedish communities, this definition is accompanied by the explanatory subordinate clauses '*que chacun peut améliorer*' [that everyone can improve], '*que todos pueden editar*' [that everyone can edit], '*som alla kan redigera*' [that everyone can edit] ('Wikipédia:Accueil_principal' 2021; 'Wikipedia:Portada' 2021; 'Portal:Huvudsida' 2021). Another instance of shared values and related enterprises is found in the already mentioned 'Five pillars'. These core principles, reproduced here again for the sake of clarity, state that a) Wikipedia is an encyclopaedia; b) Wikipedia is written from a neutral point of view; c) Wikipedia is free content that anyone can use, edit and distribute; d) Wikipedia's editors should treat each other with respect and civility; and e) Wikipedia has no firm rules ('Wikipedia:Five Pillars' 2021).

As the next subsections will show, these commonalities go beyond overarching values such as accessibility, civility and editorial neutrality. While these examples

provide sufficient grounds to classify Wikipedia as a constellation, there are other universal principles shared by more specific documented standards on translation. In the remainder of this chapter, these principles or values are discussed in relation to translation practices across the four Wikipedia language communities. In particular, the following sections will bring to the forefront some of the most common standards of practice that underpin translation in Wikipedia.

### 4.1.1 Verifiability of the ST

All four Wikipedia communities place emphasis on the verifiability of the sources. This is hardly surprising, given that 'Verifiability' (VER) is an official core policy of the user-generated encyclopaedia. VER states that the information provided in a Wikipedia article must come from a reliable source, even when the editor is 'sure that something is true' ('Wikipedia:Verifiability' 2021). Consequently, the translation guidelines of the four Wikipedia communities draw heavily on this principle and stress the importance of certifying the authenticity of the references listed in the ST. For instance, the Spanish Wikipedia translation guidelines recommend avoiding 'unverified information and articles displaying the maintenance template'[20] ('Ayuda:Cómo traducir un artículo' 2020), whereas the French Wikipedia guidelines inform potential translators that 'the facts presented in the [source] article must be verifiable by sources and references' [21] ('Aide:Traduction' 2020).

Similarly, the Dutch and Swedish Wikipedia translation guidelines are quite explicit about VER. In the Dutch Wikipedia, editors are discouraged from translating articles 'in need of citations', as doing so would transfer the problem to the target language ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). The Swedish Wikipedia essay[22] tackles the verifiability problem by alluding

[20] A template is a script that uses parsing functions. It contains repetitive material that features in various Wikipedia pages. The encyclopaedia utilises a wide range of templates, mostly for maintenance purposes. For instance, templates can indicate that an article is either missing information or is not meeting certain Wikipedia rules. Administrators, other editors and bots may also use templates to warn disruptive users against the consequences of violating policies ('vandalism').

[21] The quotations that appear throughout this chapter come from the translations of the documents.

[22] As explained in 3.2.1, the Swedish Wikipedia essay *Översättningsrekommendationer* [Translation recommendations/advice] examined in this study was renamed on 16th July 2021. At the time of

to the English Wikipedia, since this is the most frequent source of cross-wiki translations. Thus, Swedish Wikipedia editors are reminded that although the English Wikipedia may be more comprehensive, there is no reason to believe that the information contained there must be trusted blindly. The text uses a cooking analogy to illustrate this point: 'Wikipedia articles, especially the longer ones, are cooked by many different chefs, some of whom may have motivations that have little to do with writing a neutral and verifiable encyclopaedia'. Therefore, when possible, translators into the Swedish Wikipedia should attempt to avoid the rendering of unsubstantiated facts ('Wikipedia:Översättningsrekommendationer' 2020).

### 4.1.2 Acknowledgement of the source(s)

Closely related to the previous recommendation, acknowledging the sources is another value in which the four Wikipedia communities seem to agree. The Spanish Wikipedia guidelines ('Ayuda:Cómo traducir un artículo' 2020) have a section devoted to copyright wherein three points are highlighted: a) that neither Wikimedia Foundation nor Wikipedia own the copyright of the articles; b) that all Wikipedia articles belong to the editors involved in their creation; and c) that translators are free to use and reproduce Wikipedia content if they acknowledge the source(s). The French Wikipedia guidelines ('Aide:Traduction' 2020) state that while 'Wikipedia's license allows everyone to translate from one language community to another', it is obligatory to acknowledge the original author(s) and keep all the references that were originally part of the ST.

Within the same Wikipedia community, another guideline ('Wikipédia:Traduction automatique' 2018) stresses the importance of respecting Wikipedia's license and warns editors against copyright violations. Under the heading '*Conséquences*' [Consequences], editors are advised to be careful when translating content from external sources since such content is likely to be protected by copyright. According to the same page, non-compliant editors will have their articles systematically removed from Wikipedia pursuant to the criteria for deletion.

writing, its name is *Rekommendationer vid översättning från engelska* [Recommendations for translating from English].

In the Dutch Wikipedia guidelines, editors receive similar advice concerning how to acknowledge the sources. Nonetheless, there are some marked differences when it comes to the criteria for citations. Unlike the French Wikipedia, translators into Dutch are discouraged from citing sources they have not consulted, even when they are 'referenced in the original article' ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). Regarding license, editors learn that, in compliance with Wikipedia's terms, it is crucial to 'mention which Wikipedia article served as a source' for the translation. Despite the citation caveat, this section of the guidelines resembles the French and Spanish Wikipedia documents in tone and layout, thus suggesting that it is quite likely that the three Wikipedia language communities are drawing on material from the same source. For example, the remainder of the Dutch Wikipedia section on license agreements stipulates that 'everyone is free to use texts [from another Wikipedia] on condition that the source and its creators are acknowledged and that the resulting work is published under the same license' ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019).

The Swedish Wikipedia essay is ambiguous about license. However, despite not being mentioned in the text, the Swedish language community is still bound by the same license agreements that apply to all Wikipedia communities. This becomes apparent when one consults the Swedish version of the overarching Wikipedia policy on copyright ('Wikipedia:Upphovsrätt' 2020), which further clarifies the license terms.

### 4.1.3 Automatic translation

Alongside verifiability and licensing, the four communities coincide on their categorical rejection of unrevised automatic translations. In the Spanish Wikipedia guidelines, the text leaves no doubt: 'automatic translations are not allowed' ('Ayuda:Cómo traducir un artículo' 2020). Proof that the matter is taken seriously is found in the French Wikipedia, which has guidelines dealing exclusively with automatic translation ('Wikipédia:Traduction automatique' 2018). Throughout the text, translators into French are reminded of the problems arising from machine translation: a) it often results in incomprehensible and inaccurate texts; b) it is a step backwards when compared to manual translation because machine-generated texts

often require a thorough revision; and c) its use is potentially harmful to Wikipedia if editors are not well versed in the handling of the devices. As 4.3.2 will show, opinions within the French Wikipedia community are divided regarding the deployment of automation.

The Dutch Wikipedia establishes that automated content-creation devices can be mobilised to create articles only if editors make sure that the TT is revised thoroughly. The responsibility lies with the translators, who are expected to use the devices critically. This is enunciated in the text, where editors are forewarned of the outcome should they fail to follow the standards: 'unrevised machine-generated texts will be deleted without exception' ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). Likewise, the Swedish Wikipedia essay suggests the same course of action. Thus, translators are discouraged from using automatic translation, 'not even as the basis for manual translation' ('Wikipedia:Översättningsrekommendationer' 2020).

### 4.1.4 Linguistic prescriptivism

Another point of convergence concerns clarity of language. Despite having different approaches to style, the four Wikipedia communities lean toward linguistic prescriptivism. The Spanish and French Wikipedia communities state that the translated article must be perfectly comprehensible in their respective languages, and they both indicate preference for a standard international variety, void of regionalisms ('Ayuda:Cómo traducir un artículo' 2020), and easy to follow by all readers regardless of their origin ('Aide:Traduction' 2020). The two communities also encourage editors to be watchful of false friends and, in the case of the Spanish Wikipedia, there are even some pointers on how to avoid importing Anglicisms.

The Dutch and Swedish Wikipedia communities provide more detail about common language issues encountered by translators. Unlike their Spanish and French counterparts, however, these two communities have compiled an extensive manual of style aimed at translators. The Dutch Wikipedia translation guidelines warn editors against not only false friends, but also barbarisms and culture-specific terms ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). By applying the motto 'a good translator is a proficient writer in Dutch',

these prescriptive guidelines give advice on how to ensure or enhance the quality of the translations: 'translate the text, leave it aside for a few days, then read it again without looking at the original; if you think you would express yourself differently in Dutch, rewrite your sentence' ('Help:Tips' 2019).

The Swedish Wikipedia translation essay is more exhaustive. Notwithstanding, there is some overlap or continuity between the information displayed on the page and that found in the other three language communities. Those parts of the text that diverge will be analysed in 4.2. For now, suffice it to say that in a manner similar to the Dutch Wikipedia guidelines, the Swedish Wikipedia essay asks editors to 'read the [translated] sentence aloud many times and modify it until it sounds natural [to them]' ('Wikipedia:Översättningsrekommendationer' 2020). The same text encourages editors to 'use the linguistic mechanisms available in Swedish', a language that 'has a relatively more flexible syntactic structure than English'.

### 4.1.5 Grammar and style

As was observed above, linguistic prescriptivism is not uncommon in Wikipedia. In the Spanish Wikipedia, translators are expected to avoid Anglicisms and choose vernacular Spanish lexical items. In the French Wikipedia, Anglicisms are not frowned upon, but the guidelines recommend being careful with false friends ('Aide:Traduction' 2020). In the Dutch and Swedish Wikipedia communities, editors are also encouraged to keep an eye on possible false friends, but the documents go into more detail about other possible language issues that may arise during the translation process. In the Dutch Wikipedia, translators are informed that other Wikipedia communities have different conventions for representing homonyms ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). For example, while the English Wikipedia uses a comma, other languages – including Spanish and French – utilise brackets to disambiguate terms. Thus, according to *Tips*, the Wikipedia article for the Maltese town of Rabat would be [[Rabat, Malta]] in English, but [[Rabat (Malta)]] in Dutch and languages following similar conventions. In other words, translated articles are subject to the naming conventions set by the target Wikipedia community. Another aspect highlighted by the Dutch translation guidelines has to do with the use of external links in the running text. Again, while this seems to occur in the English Wikipedia,

it is considered bad practice in the Dutch Wikipedia, where their inclusion is forbidden.

More related to grammar than to style, the Swedish Wikipedia offers a series of tips on how to avoid common transfer errors. Translators into Swedish have a manual of style aimed primarily at peripheral members of the community (Wenger 1998; Shove, Pantzar, and Watson 2012), that is, editors with little to no experience in translation. Besides the points already mentioned, the Swedish essay covers diverse aspects of syntax, semantics, and systems of measurement. As regards syntax, the text provides a simple, yet illustrative example of how English and Swedish use different sentence structures. Thus, the document explains that the former tends to rely on commas to separate lexical items in a sentence, including cases of enumeration, whereas Swedish prefers shorter sentences and no comma after the second last option in an enumeration. Therefore, translators are expected to be careful and write '*ett, två, tre och fyra*' [one, two, three and four] without a serial comma between the last two items, as would happen in the English Wikipedia ('Wikipedia:Översättningsrekommendationer' 2020).

These errors are also found in lexical choices. The guidelines recommend 'watch[ing] out for musical notes above A' when translating music articles from the English Wikipedia. Since both languages have traditionally used different nomenclature for musical notes, Swedish Wikipedia translators should revise the text and change those notes to suit the standards established by the target community. Similar advice is given regarding systems of measurement such as imperial versus metric, the former being widely used in English-speaking countries, especially in the United States. Other tips are related to prepositions, transliteration of certain names, passive constructions, capitalisation, and genitives in Swedish ('Wikipedia:Översättningsrekommendationer' 2020).

### 4.1.6 Content Translation Tool (CX)

As discussed in 1.1.2, the Spanish and French Wikipedia communities feature as the two most prominent users of CX, with ca. 92,000 and 73,000 translated articles, respectively. These figures are in stark contrast with the number of CX-generated articles in the Dutch (ca. 7,100) and Swedish (ca. 3,100) Wikipedia communities

(‘Content Translation Statistics’ 2021). Although these significant differences will be addressed in 4.3, it is worth noting that all four Wikipedia communities have equal access to the device. The French and Dutch communities have a page devoted to CX. Their Spanish and Swedish counterparts do not have a dedicated page, but editors of these communities can still access the user guide in their languages on MediaWiki, a free and open-source wiki engine developed for Wikipedia in 2002 (‘Content Translation’ 2020).

The platform serves as a bridge between Wikimedia projects, including Wikipedia. As CX is aimed at facilitating translations across different language communities of Wikipedia, its multilingual instruction manuals, irrespective of their location in MediaWiki, are addressed to editors of the user-driven encyclopaedia. In reading the user guide in both Meta and the Dutch and French Wikipedia communities (‘Wikipedia:ContentTranslation’ 2018; ‘Aide:Outil de traduction’ 2020) , it is possible to identify the English version of the text in MediaWiki as the ST (‘Content Translation’ 2020). Translators of the four communities have rendered the text into their languages, maintaining certain uniformity. For instance, the primary set of instructions remains virtually unchanged across languages. The steps concerning how to activate the tool, select an article for translation and work through the text are reproduced in the different versions.

The four language versions also recommend using the device critically, especially since ‘[it] is not fully operational and using it may come with bugs or surprises’ (‘Wikipedia:ContentTranslation’ 2018; ‘Aide:Outil de traduction’ 2020; ‘Content Translation’ 2020). Although these bugs are mentioned in the four languages and may cast doubts on the functionality of CX, there is a clear gap between widely spoken languages such as Spanish and French, and languages like Dutch and Swedish. This asymmetry does not go unnoticed in the Dutch Wikipedia, where editors are informed that Wikimedia Foundation is working alongside free software developers ‘to provide open translation software for Dutch and a number of other languages’ (‘Wikipedia:ContentTranslation’ 2018). The same text clarifies that for languages such as Spanish and Portuguese ‘the software is already available’.

The gap between major and minor languages may explain why the Spanish and French translations of the English text in MediaWiki are more optimistic about

Wikipedia's bespoke device. In the second paragraph, readers learn that 'while still in active development, [CX] is available for you to try and has been used already to create thousands of articles resulting in reported improvements in translators' productivity' ('Content Translation' 2020). This optimism about the device is omitted in the Swedish translation of the text: '[the tool] has already been used to create thousands of articles faster than in the traditional way' ('Content Translation' 2020).

## 4.2 Differences in the standards of practice across Wikipedia communities

Having examined the similarities between the standards of the four Wikipedia communities, this section turns the attention to the divergences. As was mentioned in 2.2, three defining features of a community of practice are the existence of a shared repertoire, a joint enterprise, and processes of mutual engagement (Wenger 1998; Farnsworth, Kleanthous, and Wenger-Trayner 2016). The idea that different individuals engage with one another in the performance of a series of routinized distinctive activities is recurrent in the practice theory literature (Nicolini 2012; Shove, Pantzar, and Watson 2012; Buch and Schatzki 2019), even among scholars who focus on practice as a conceptual or analytical unit rather than as the property of a community. Therefore, an indispensable element of practice is the willingness of a group of people – practitioners – to engage and perform their tasks following certain standards and procedures that are not always clear to outsiders (Wenger 1998; Warde 2005; 2016). Thus, what distinguishes one community from another are differences in standards and procedures. In Wikipedia, despite the already seen elements of continuity, each language community has some leeway to set its own standards. In the next subsections, this set of values and expectations that underpin local translation practices are examined in more detail.

### 4.2.1 Selecting a Wikipedia article for translation

Although values such as quality and verifiability are common across language communities and can be regarded as the cornerstone of the online encyclopaedia, there are a few underlying cross-wiki differences in how these criteria are met. These divergences are found, for instance, in the criteria for selecting an article for

translation. The Spanish Wikipedia guidelines show a clear preference for featured articles[23] (FA) in other languages as the point of departure in the translation process. In the text, editors are encouraged to 'translate featured articles because these meet all the quality criteria established by Wikipedia' ('Ayuda:Cómo traducir un artículo' 2020). As a second option, the same guidelines recommend translating high-quality articles and paying attention to 'what other editors have said about their content' on the talk pages ('Ayuda:Cómo traducir un artículo' 2020). According to the text, if editors have raised important concerns about the reliability or neutrality of the source article, then it should not be translated. Moreover, the guidelines advise strongly against translating articles from other Wikipedia communities if those articles are 'neither featured nor high quality'.

Considering that only a reduced number of Wikipedia articles have been awarded that distinction, the selection criteria seem to be quite restrictive. In the English Wikipedia, by far the most common source of Wikipedia translations via CX, there are over six million articles. Of these, fewer than 6,000 are featured. This means that less than 0.1% of the total number of articles meet the 'ideal' criteria for translation set by the Spanish Wikipedia guidelines. While these are simply recommendations and editors are still allowed to translate other Wikipedia articles as and when they please, it is worth noting the complexities of rendering FAs into any language. This type of article tends to be much longer than the average Wikipedia article: FAs are meant to be exhaustive, drawing on information from multiple reliable sources. Consequently, they may pose significant challenges to an inexperienced translator, who may find the task overwhelming. It is perhaps for this reason that the French Wikipedia guidelines have more flexible criteria, suggesting the translation of 'articles of a certain quality' which, in any case, must be 'chosen carefully', paying attention to notability [24], verifiability and neutrality ('Aide:Traduction' 2020).

[23] In Wikipedia, featured articles (FA) are articles that have been subject to a thorough revision process by the community and have been awarded a 'badge' (usually in the form of a star featuring on the top right) on the basis of meeting the highest standards of quality. Anyone with a registered account can nominate an article for this distinction if they consider that the article is well written and meets all the essential criteria, especially verifiability and neutrality.

[24] According to Wikipedia, 'notability is a test used by editors to decide whether a given topic warrants its own article' ('Wikipedia:Notability' 2021). Discrepancies often arise between editors and across language communities as to what counts as relevant and worthy of being in Wikipedia.

The Dutch and Swedish Wikipedia communities place the emphasis on quality as the primary criterion, and the advice is geared towards ensuring that translators possess some knowledge about the topic covered in the source article. The Dutch Wikipedia translation guidelines, in particular, indicate that familiarity with the subject is not just desirable but mandatory. According to the text, editors should '*only*[25] translate topics [they] are familiar with' ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). There are also suggestions of what not to translate. The text goes on to say that 'several texts have a lot of detail, mainly concerning local affairs' that may not be relevant to a Dutch Wikipedia reader and can therefore be summarised or omitted. The same issue is brought forward in the Swedish Wikipedia essay, which states that some English Wikipedia articles have 'far too much information to be really effective as encyclopaedic texts' ('Wikipedia:Översättningsrekommendationer' 2020).

It is also worth noticing that the Swedish Wikipedia documented standard diverges from those of the other communities under investigation in that the duties of the translators are more clearly formulated in the text. Apart from summarising content, editors are asked to use their common sense: 'a translator's most important characteristic is to understand when [they] have not understood' ('Wikipedia:Översättningsrekommendationer' 2020). The text highlights the importance of critical thinking and doing research when some concepts are particularly difficult to translate. In addition, editors are strongly encouraged to contact the author(s) of the ST if they are still in doubt after searching for the correct terminology. Another feature of the Swedish Wikipedia essay is that, in some cases, it resorts to irony. For instance, editors are reminded that Wikipedia is available in many other languages beside the 'super[ordinate] language English'. Thus, Swedish Wikipedia editors are 'emboldened' to translate from other Scandinavian languages such as Danish and Norwegian.

[25] My emphasis.

### 4.2.2 Use of translation templates

As far as technical knowledge is concerned, both the Spanish and French Wikipedia translation guidelines provide editors with a series of bespoke templates. Such templates should be added to the article to indicate that the translation is either in progress or exhibits considerable errors. The objective of these templates is to warn readers and editors alike that the article may contain significant shortcomings that have to be addressed. In the case of ongoing translations, the template also serves an important purpose: it shows that one or more editors are working on the text and that they are aware of its inaccuracies. In other words, adding a template to an article that is being translated gives editors a safe passage; it informs Wikipedia administrators that the article should not be deleted. This guarantee, however, also entails responsibilities: editors must agree to improve the article and remove the template only when they consider that the issues have been resolved.

The selection of a suitable translation template for the article is somewhat subjective since it depends on an editor's interpretation and enforcement of Wikipedia's policies. The Spanish Wikipedia guidelines ('Ayuda:Cómo traducir un artículo' 2020) contemplate four possible outcomes. According to the text, if the translation is clear but contains significant errors, editors should use the 'revise translation' template; if, on the contrary, the translation is unclear or incomplete, they ought to apply the 'bad translation' template. In cases where the translation is automatic and has not yet been revised, the document recommends using the 'automatic translation' template. If the text is written in another language or the translation is incomprehensible, a sensible course of action suggested in the text is to add the 'destroy' template to the article and wait for an administrator to delete the page ('Ayuda:Cómo traducir un artículo' 2020).

The French Wikipedia guidelines ('Aide:Traduction' 2020) offer a concise summary of the main steps to be considered prior to publishing the translation. First, editors that choose to translate manually instead of using CX are advised to 'copy the wikicode of the source article and paste it into [their] user subpage'. Secondly, upon revising the translation, editors are encouraged to move their draft to Wikipedia's 'main space'. Moving the text to the main space means that the article becomes part of the encyclopaedia and can thus be traced using the platform's

search engine. Thirdly, beginners are reminded that, should they wish to do so, they can contact more experienced editors of the community who 'will help [them] improve the article and teach [them] more about the methodology' ('Aide:Traduction' 2020). Finally, both the French and Spanish Wikipedia guidelines recommend linking the [translated] article to Wikidata. This last step, as explained in 2.4.1, is crucial for the article to be connected to other Wikipedia language versions in which similar content is available. A successful linkage to Wikidata will show a list of languages on the left sidebar of the article.

Thus far, the analysis has shown that all the elements that make up the guidelines have clear addressees: editors with little to no prior experience (novices or newcomers) in cross-wiki translation. The guidelines are intended to initiate peripheral members and facilitate their successful inbound trajectory to membership (Wenger 1998). To meet this end, each text contains examples of what constitutes a good practice. Arguably, and as will be investigated in 4.3, these examples result from processes of mutual engagement whereby more experienced Wikipedia editors come together and decide on what will work best for their communities. Thereafter, they seek ways to transmit those decisions to novitiates by making sure that they have access to standards of performance. As noted by Warde (2005, 138), most practices have longstanding members who are tasked with providing beginners with the knowledge and tools that are necessary for their correct fulfilment. In the case of Wikipedia, experienced editors can also impose a series of restrictions on who has access to specific devices. The next subsection tackles this issue by bringing to the forefront the mobilisation of bots for content-creation purposes.

### 4.2.3 Bots

As explained in 1.1.2, bots are an integral part of Wikipedia (Geiger 2011; Geiger and Halfaker 2017). Introduced in 2002, these automated devices have been in the encyclopaedia almost since its creation and are run by trusted members of the community with some basic knowledge of programming languages ('Help:Creating a Bot' 2021). Bots are used for multiple purposes including, but not limited to, the correction of spelling mistakes, the early detection and reversion of blatant vandalism and copyright infringements, and the archiving of long idle threads on

talk pages. Additionally, bots warn disruptive users or 'vandals' of the consequences of their wrongdoings, and inform administrators of possible violations of Wikipedia's policies. In doing so, bots use algorithms that are constantly being updated to ensure maximum efficiency (O'Hagan 2016; Tsvetkova et al. 2017).

In 2.4, it was argued that bots and translation devices are competing, yet complementary, means of article creation. However, unlike bespoke Wikipedia translation devices such as CX, bots have been the subject of much debate among editors. Despite playing a key role in the encyclopaedia and facilitating the performance of repetitive, mundane and time-consuming tasks (Lih 2009; O'Hagan 2016), notable discrepancies arise regarding their use in the creation of encyclopaedic articles. On the one hand, some Wikipedia language communities, in particular those with a reduced number of active users, have opted to deploy bots more freely. On the other side of the spectrum, Wikipedia communities with larger numbers of editors are known to have restricted their mobilisation. These restrictions mean that only senior editors who fulfil certain criteria are allowed to run bots and, even when permission is granted, it is always temporary and under close scrutiny ('Wikipedia:Creación de artículos con bot' 2020).

As Table 4-1 below illustrates, there is a tendency among minor national and regional languages to rely on bot usage to bridge the gap between readers and editors. The prevailing lack of active editors in a small Wikipedia community results in a fairly limited number of articles. In other cases, it is not the number of editors that leads communities to choose bots but, rather, the need to create a large number of articles on a specific topic within a short period. Additionally, Table 4-1 reveals that, on the surface, Wikipedia communities with a larger number of editors are less dependent on bots for article creation. Chapter 6 will investigate these differences further by analysing data obtained from the interviews with senior Wikipedia translators.

| Bot-generated content per Wikipedia community | | | | | |
|---|---|---|---|---|---|
| **Restricted use of the tool** | | | **Active use of the tool** | | |
| **Community** | % of bot-generated articles | Number of bot-generated articles (Total number of articles) | **Community** | % of bot-generated articles | Number of bot-generated articles (Total number of articles) |
| German | 0% | 1.2K (2.3M) | Cebuano | 100% | 5.4M (5.4M) |
| Greek | 0% | 675 (157K) | Malagasy | 94% | 85K (91K) |
| Japanese | 0% | 140 (1.1M) | Uzbek | 91% | 118K (130K) |
| **Spanish** | **0%** | **2.2K (1.5M)** | Waray-Waray | 90% | 1.1M (1.3M) |
| Czech | 1% | 3.0K (419K) | Volapuk | 89% | 109K (122K) |
| Korean | 1% | 4.5K (439K) | **Swedish** | **81%** | **3.1M (3.8M)** |
| English | 3% | 150K (5.8M) | Asturian | 74% | 74K (100K) |
| **French** | **4%** | **79K (2.1M)** | Serbian | 60% | 368K (614K) |
| Italian | 7% | 110K (1.5M) | Basque | 56% | 172K (310K) |
| Russian | 10% | 149K (1.5M) | **Dutch** | **54%** | **1.0M (2.0M)** |

Table 4-1. Bot generated-articles per Wikipedia language community. K stands for thousand, M for million.

When it comes to the mobilisation of bots, a vast number of Wikipedia communities have official policies regulating their functioning. This is due to the variety of tasks these automated devices can perform and the potential disruption their failure may cause. Nonetheless, not all Wikipedia communities have the same regulations. For example, the Dutch Wikipedia guidelines ('Help:Gebruik van bots' 2020) state that 'there are only few rules for bot usage' in that community. This statement is

followed by a piece of advice to editors wanting to run a bot in other Wikipedia communities: 'Watch out when you use bots on Wikipedias in other languages, for example the English one, [since] other rules apply'.

In the Spanish Wikipedia, besides the official bot policy found in most Wikipedia communities, there is a policy aimed at overseeing bot-generated content. The policy ('Wikipedia:Creación de artículos con bot' 2020) was drafted on 1st April 2007 by an administrator with the objective of 'imposing certain restrictions, ensuring that no abuses will be committed that can put the quality of the Spanish Wikipedia at risk'. The text was approved on 20th January 2008 by the Wikipedia community, with 90 votes for and 40 against ('Votaciones:Botopedia' 2008). The template at the top of the page provides information regarding its scope and enforcement: '[The policy] has been compiled and approved by the community and its compliance is obligatory for all editors' ('Wikipedia:Creación de artículos con bot' 2020). This kind of engagement, its mechanisms and its significance will be examined further across in 4.3.1-4.3.4.

Following the approval of this policy, bot-generated articles became increasingly rare in the Spanish Wikipedia. This was shown in Table 4-1 above, where the percentage of articles created by a bot was rounded to zero as of July 2019, when the statistics were last updated ('Wikipedia Statistics' 2019). This means that the number of bot-generated articles in the community is below 2,200. Such small figures are likely to be the outcome of stringent rules dictated by the policy. Thus, those who wish to run a bot for content-creation purposes must meet the following criteria: a) be trusted members of the community; b) have a well-defined project that justifies bot usage; and c) specify in their application if the [bot-generated] articles will be expanded or modified by humans once they have been created ('Wikipedia:Creación de artículos con bot' 2020).

Although the concept of trust can seem subjective at first glance, the text narrows it down to editors that show familiarity with the shared repertoire. Specifically, the document states that editors must be able to 'prove their knowledge of Wikipedia policies, have a background in programming language, and [be] versed in the topic they intend to use the bot for' ('Wikipedia:Creación de artículos con bot' 2020). Moreover, in submitting the application for bot usage, there is a 14-day open voting

process in which members of the community that endorse or reject the project are encouraged to justify their decision. Each application is considered on a case-by-case basis, taking into account not only the feasibility of the proposal but also the applicant's trustworthiness, assessed by their positive record of contributions to the community. Whether or not a proposal is suitable depends on the applicant being able to set realistic goals and justify the need for bot usage in the creation of an established number of articles. One example would be to fill an important knowledge gap. Permission is granted temporarily, usually until the aims of the project have been achieved, and only when the proposal receives the support of two thirds of the voters. Articles created after the permission has expired are deleted. Likewise, bot-generated articles will be considered for deletion if they exceed the number specified in the proposal or if they have not been revised after some time has elapsed ('Wikipedia:Creación de artículos con bot' 2020).

The French Wikipedia, despite not having an official policy on bot usage for creating articles, still imposes similar restrictions on who can run these automated devices. In the policy ('Wikipédia:Créer un bot' 2020) [Creating a bot], it is stated that in order to use a bot some 'previous knowledge of programming languages is required'. As in the Spanish Wikipedia, permission is granted on a case-by-case basis, and each applicant must specify in the bot user page what the intended goals are. The community then has to decide whether to grant permission. The device remains under constant supervision and authorisation can be suspended either temporarily or permanently if the bot is misused or is not working properly ('Wikipédia:Créer un bot' 2020). As shown in Table 4-1, the number of bot-generated articles in the French Wikipedia is, at the time of writing, close to 79,000, which is significantly higher than those of the Spanish Wikipedia. However, they still represent less than 4% of the 2.1 million entries available in that community.

The Dutch and Swedish Wikipedia communities differ radically from their Spanish and French counterparts in how they manipulate bots for creating articles. As Table 4-1 above shows, the Dutch and Swedish Wikipedia communities rely heavily on the use of automation, with 54% and 81% of the total number of encyclopaedic entries being created by bots, respectively. These percentages indicate remarkable differences in the mobilisation of materials between the four communities. As a result of bot usage, the Dutch Wikipedia contains two million entries, and the

Swedish Wikipedia surpasses 3.8 million. Chapter 6 will investigate the reasons underpinning bot deployment in more detail. Among other things, the interviews aim to shed light on what prompted Dutch and Swedish Wikipedia communities to incorporate materials such as bots into their practices.

Neither the Dutch nor the Swedish Wikipedia communities have strict rules in place to regulate bot-generated content. This fact became apparent after analysing the Dutch Wikipedia guidelines ('Help:Gebruik van bots' 2020), where, as mentioned earlier, the text states that 'there are only few rules for bot usage'. Notwithstanding, it would be careless to assume that bot-generated content in these communities is uncharted territory. In fact, the Dutch Wikipedia page on bots has a brief section entitled '*Botmatige aanmaak van nieuwe artikelen*' [Creating new articles with bots], where editors are encouraged to notify other members of the community if they intend to create articles using the device. The procedure is more a recommendation than a command, a gesture of courtesy due to the fact that unreported large-scale bot creations can cause controversy in the community. This is because bot scripts contain too much data and they can slow down the site ('Help:Gebruik van bots' 2020).

The Swedish Wikipedia guidelines ('Wikipedia:Robotar' 2019) [Robots] provide a succinct user-guide explaining what bots are, what they do and how to run them. Unlike the Dutch Wikipedia, which has a short, albeit informative, section on bot-generated content, the Swedish Wikipedia page provides an account of bots used in maintenance tasks. Throughout the text, editors are instructed to follow the steps closely and submit an application to be considered by the community. In the application, editors are advised to specify the intended goals for which bot usage is required. This request is followed by a warning that 'bots cannot be used to make controversial edits', and that a misuse of the automated device will result in administrators stepping in and taking action ('Wikipedia:Robotar' 2019). Moreover, bots can be blocked temporarily by an administrator if the owner does not abide by the rules or if the device is malfunctioning.

As this first stage of the analysis has shown, Wikipedia can be regarded as a constellation of communities of practice sustained by shared values, similar goals, and a series of overlapping policies and guidelines. At the same time, each

Wikipedia community has a separate shared repertoire consisting of more localised standards either dictating or giving advice on how to translate and mobilise content-creation devices. Thus, while some policies such as those featuring in the 'Five pillars' apply to all communities, it is up to each language community to set up their own rules and ensure their enforcement. For instance, the Spanish Wikipedia has stricter controls on who can run bots and how to obtain the community's approval, whereas other communities such as the Dutch and Swedish ones are characterised by a relative laxity of standards for bot-generated content. Yet, as became clear in this last subsection, there are occasional pockets of resistance within the communities. The Spanish policy ('Wikipedia:Creación de artículos con bot' 2020) received 40 votes against its approval, and even communities with more flexible rules such as the Dutch Wikipedia recognise that misreported large-scale bot creations can generate unrest among some members ('Help:Gebruik van bots' 2020). In the next stage of the analysis, which targets comments posted by editors on talk pages, the internal processes of mutual engagement and negotiation, including rebellion, are scrutinised.

## 4.3 Negotiating translation standards on talk pages

This second stage of the analysis investigates how the selected documented standards are negotiated in each community. As discussed in 3.3, the investigation draws on data from comments posted on talk pages wherein editors discuss and propose amendments to the texts. The aim is to gain insight not only into the volatility of translation practices but also into how editors engage with one another and decide on what is best for their local Wikipedia communities. Thus, prominence is given to discussions that have had a perceptible impact on the translation standards examined in 4.1 and 4.2. Since the ultimate goal is to understand how individual Wikipedia communities negotiate their translation standards, the findings are grouped by community. In keeping with the order followed in the preceding analytical phase, the first two subsections will focus on the Spanish and French Wikipedia communities, and the last two will concentrate on their Dutch and Swedish counterparts.

Despite the diversity of issues subject to Wikipedia negotiations, the process is relatively straightforward. An editor initiates a thread on the talk page, which

contains a title, the body of the text, and a signature at the end of the message indicating the name of the author and the date when the comment was posted (see 3.3.1). On reading the posting, any member of the community is free to engage in the debate, either by replying within the same thread or by commencing a new one. Most threads comment on aspects related to a specific Wikipedia standard or article. As Wenger, McDermott, and Snyder (2002) posit, forum-like spaces are relatively common in distributed communities of practice, where members are geographically dispersed and cannot rely on face-to-face communication. These forums contribute greatly to developing a 'strong sense of craft intimacy' around common concerns, where practitioners engage with one another and work jointly to find solutions (Wenger, McDermott, and Snyder 2002).

As will be discussed in the next subsections, the degree of participation and engagement on talk pages varies considerably within and across communities depending on the nature of the subject. A successful Wikipedia negotiation typically involves one or more editors putting forward changes to a page in one of the threads. These proposals are then considered by other members of the community, usually with a robust knowledge of Wikipedia's regulations. In the case of policymaking, where a major consensus is needed (see 2.2.1), editors engage in both talk page discussions and voting. In what follows, these different avenues of negotiation are examined in detail.

### 4.3.1 Negotiating bots and sources in the Spanish Wikipedia

This section examines two different negotiation processes: policymaking and amendments to the guidelines. The first one originates on the talk page linked to the aforementioned bot-generated content poll and primarily involves senior members of the community. The second one, displaying a more diverse group of editors, revolves around issues of verifiability and acknowledgement of the sources and occurs on the talk pages associated with the translation guidelines. As this section will investigate, whilst overarching values have received some attention, Spanish Wikipedia editors are visibly more engaged in those elements that are part of their shared repertoire. In other words, those matters that affect them directly and they have some control over are normally the ones that register higher levels of

engagement. This is true of the mobilisation of bots for the creation of encyclopaedic articles.

In 4.2.3, it was noted that the official policy was approved with 90 votes for and 40 against. Although the core aspects of the policy were drafted in 2007, voting took place between 6th and 20th January 2008. The poll was announced in the *Café*, the Spanish Wikipedia 'village pump' (see 3.4.1). Data retrieved from the other 11 polls held in 2008 in the community show that this poll attracted a much higher degree of participation than other polls held during that year ('Wikipedia:Votaciones finalizadas' 2008). Of the 11 polls, only one of them registered more than 100 participants (<110). The average number of participants was 79.1. As Table 4-2 below illustrates, during the 14-day period that the bot-creation poll remained open, 45 discussion threads were initiated on the policy's talk page ('Wikipedia discusión:Votaciones/2008/Creación de artículos con bots' 2008).

In light of the large amount of data encountered while conducting the analysis, it was necessary to filter out those postings that did not address the issues raised by RQ2, reproduced here for the sake of convenience: 'How are translation standards negotiated within the four language communities?' In keeping with the selection criteria introduced in 3.3.1, only individual postings that fell within the scope of RQ2 were selected for the sample. This excluded single-authored threads containing redundant arguments, threads in which no negotiation occurred, and multiple lengthy essay-like threads initiated by the same author where the same standpoint was expressed. For illustrative purposes, Table 4-2 below presents a general list of themes discussed throughout the 45 threads. Those threads that were selected for the analysis are marked in bold. As most threads involve at least two participants who do not necessarily share the same views, the thematic classification was done taking into consideration the argument presented by the editor that initiated the corresponding thread.

| Theme | Thread(s) |
|---|---|
| Reminders, in general: reading the text carefully before voting, specifying deadlines, asking for reflection, announcing results | 1.1, **1.5**, 1.9, 1.22, 1.32, 1.44 |
| Bot-generated article quality (length, information, manual expansion once created, etc.) | 1.2, 1.3, 1.26 |
| Expressing dissent (reasons for voting against the policy, scepticism about bots) | **1.4**, **1.8**, 1.10, **1.11**, **1.12**, 1.15, 1.16, 1.18, 1.20, 1.24, 1.25, 1.27, 1.28, 1.29, 1.31, 1.33, 1.34, 1.35, 1.36, 1.37, 1.38, 1.40, 1.41, **1.45** |
| Permissions (use of license/flag) | 1.6 |
| Negotiation: trying to convince others to vote for the policy | **1.7,** 1.13**,** 1.14**,** 1.17, 1.42 |
| Suggesting minor changes to the text | 1.19, 1.21, 1.23, 1.43 |
| Expressing neutrality | 1.14 |
| Accusing editors of monopolising the debate | 1.30, 1.39 |

Table 4-2. Thematic summary of all the discussion threads initiated between 6th and 31st January 2008.

The editors that participated in the discussion are listed in Table 4-3 below, alongside their date of registration in the Spanish Wikipedia[26] and their status within the community, i.e. administrator, experienced editor, novice/peripheral editor (see 2.2.1). Following the criteria set in 3.3.1, their user names were anonymised. Therefore, editors are referred to as 'ES X', where ES is the ISO

[26] As indicated in 3.3.1, the date of registration is retrieved from the editors' statistics tab and track record of contributions, both accessible through the revision history. This can be done following four basic steps: a) click on the editor's name, b) access their user page, c) revise the revision history of their page, and d) select their 'oldest contributions', which will reveal the date of the first contribution ('edit') they made as registered users.

standard two-letter abbreviation used for Spanish and X stands for the order in which editors are mentioned in the text and not their registration date.

| Editor | Year of registration in Wikipedia | Status within the community |
|---|---|---|
| ES1 | 2005 | Administrator |
| ES2 | 2005 | Administrator |
| ES3 | 2006 | Experienced editor |
| ES4 | 2005 | Administrator |
| ES5 | 2006 | Experienced editor |
| ES6 | 2005 | Administrator |
| ES7 | 2006 | Experienced editor |
| ES8 | 2001 | Administrator |
| ES9 | 2006 | Administrator |
| ES10 | 2006 | Experienced editor |
| ES11 | 2004 | Administrator |
| ES12 | 2007 | Administrator |
| ES13 | 2005 | Experienced editor |

Table 4-3. List of Spanish Wikipedia editors that participated in the debate.

The bot debate centred on quality versus quantity, the uses that these automated devices would have in the community, and the bot runner's profile ('Wikipedia discusión:Votaciones/2008/Creación de artículos con bots' 2008). Regarding quality, those who argued against the policy did so under the belief that a) bots would produce short articles with little encyclopaedic content, b) bot-generated

articles would require a thorough and time-consuming revision process, and c) only humans can write articles that are up to the standards. An illustrative example of these points is found in thread 1.4, where ES1, a Wikipedia administrator, posted that 'bots (and any machine in general) are developed to help edit articles, as in our case, and not to create articles about countries or anything else, so I will always vote against these proposals' (6th Jan 2008, 20:56 CET[27]).

One day later, in thread 1.8, the same editor extends their argument and adds that 'bots do not know how to write; [only] humans do' (7th Jan 2008 at 04:34). To support their assertion, ES1 brings to the surface their experience as a passive observer and broker (Wenger 1998) in the Portuguese Wikipedia, a community that has relied on bot usage. According to ES1, 'I have seen how bots have contributed to the impoverishment of the Portuguese Wikipedia' (7th Jan 2008 at 04:34). In the same thread, ES2 elaborates further on the Portuguese Wikipedia argument and posts:

> Bots create articles with poor content (you only have to see what happens in the Portuguese Wikipedia).
>
> The kind of measures this policy seeks to impose will take time off writing. It is better to grow not so quickly and grow better (8th Jan 2008 at 22:10).

Their statement does not go unnoticed by another senior editor that endorses the policy. On reading the previous message, ES3 explains how the implementation of bots would allow editors like them to focus on other tasks:

> I wished a bot had created hundreds of articles about Aragonese towns, so that I could dedicate time to what I am interested in, which is expanding the stubs.[28] The time I spend creating the articles prevents me from uploading hundreds of photos to [Wikimedia] Commons (8th Jan 2008 at 22:36).

As the previous posts reveal, there are two differentiated sides with clashing views on the use of bots. Each side comprises fully engaged, experienced editors that seem to be driven by what they think is best for the Spanish Wikipedia community. One of these forms of engagement, dissent, was noted by Wenger (1998), for whom

[27] Although currently the Spanish Wikipedia observes Coordinated Universal Time (UTC), back in 2008 the community used Central European Time (CET), the time zone of peninsular Spain.
[28] In Wikipedia, encyclopaedic articles with little content are known as stubs. In the Spanish Wikipedia, these articles are called *esbozos* or *microesbozos*.

disagreement between members of the same community can be regarded as a strong sign of commitment. For Wenger (1998, 77), rebellion is better than passive conformity and is a necessary driving force of change for communities of practices to evolve. Similarly, as noted in 2.3, Warde (2005, 141) postulates that 'conventions will usually be to some degree contested, with some practitioners typically still attached to prior codes of conduct'. In this case, ES1 and ES2 resist innovation whereas ES3 considers that incorporating the device into their practices would benefit the community.

These first exchanges between the two sides foreshadows that negotiation will be an arduous process. ES1, exhibiting a clearly imperialistic attitude (Wenger, McDermott, and Snyder, 2002), stresses that they 'will always vote against these proposals', and even ES2, after reading the responses given by ES3, is still reluctant to mobilise bots, adding that: 'I would rather have a bot that uploads photos to [Wikimedia] Commons than have one that creates stubs that then I have to check, delete, etc.' (9th Jan at 01:22). ES1's response is an example of what Wenger, McDermott, and Snyder (2002) have defined as factionalism in negotiation, a disorder stemming from the practitioners' joint enterprise or domain. This tendency to fight blindly for one's own beliefs and special interests from a position of strong commitment to the community's shared repertoire hinders negotiation because practitioners often take their disagreement over a particular subject to the extreme. The unrest generated by this factionalist approach becomes clear in thread 1.7 when ES4, highly involved in drafting the policy and visibly disappointed by their colleagues' anti-bot remarks, intervenes in response to ES1's initial comment: 'Damn…naturally, I feel a bit useless trying to convince someone who ends their speech saying "I am not moving from this position"' (7th Jan at 01:44).

Moving on to the role that bots would fulfil and the implications for the community, ES1, once more, expresses their discontent in thread 1.4: 'Wikipedia relies on the collaboration among volunteers, not robot volunteers' (6th Jan at 20:56). Faithful to the same factionalist strategy discussed above, ES1 believes that the Spanish Wikipedia is not heading in the right direction: 'Sadly, I am seeing how a project I helped to write manually has been taken over by machines' (6th Jan 2008 at 20:56). The editor's remarks throughout different threads prompt some immediate reaction from the pro-policy group. In thread 1.8, ES3 refutes ES1's arguments:

> I think you forget that bots are run by humans. I do not see the difference between using a bot and pressing keys on a computer keyboard. After all, a keyboard is a machine as is the computer you use (I assume). Without that machine [the computer], Wikipedia would not exist. You are going against one of the main principles of Wikipedia: that is *not written on paper* (7th Jan 2008 at 13:20).

This exchange between the two editors shows different understandings of this practice of producing Wikipedia articles. On the one hand, ES1 adopts a conservative approach, stressing that their contributions to Wikipedia have been manual. They firmly believe that creating articles manually is in accordance with the principles of the encyclopaedia, which, in their view, is exclusively about human collaboration. On the other hand, ES3 challenges that assumption by bringing to the fore the significance of materials for the existence and configuration of the practice itself. In addition, ES3 does not view bots as devices with a life of their own, capable of taking over Wikipedia, but rather, as components that facilitate the practice of editing. This approach to practice was discussed in 2.4, where, following Shove, Pantzar, and Watson (2012), and Shove (2017), it was argued that materials in general, and devices in particular, configure practices.

Besides quality and functionality, editors discuss the importance of regulating bot-generated content. In thread 1.5, another senior member of the community, ES5, observes: 'This is a turning point in the history of the Spanish Wikipedia and must not be taken lightly. I will wait a few days before casting my vote' (6th Jan 2008 at 22:20). In thread 1.8, initiated by ES1, ES6 tries to persuade sceptics with the argument that 'even without a bot policy, there have never been multiple article creations using bots' (7th Jan at 13:52). For ES6, 'regulating their usage is more positive than having loopholes' (7th Jan at 13:52). Yet, in thread 1.12, ES7 is far from convinced:

> As far as I know, such fear [of bots] does not exist. Scepticism? I do not think so. Realism? A lot and common sense. I do not believe in standardisation because it destroys creativity. If you want to create articles using bots, do so in a bot subpage, expand them, and then move them to the main space. We will then see who wants to work and who wants to inflate [the number of articles] or experiment (8th Jan 2008 at 22:48).

In response, ES8, a supporter of the policy, adds:

> An encyclopaedia gives necessary and useful information; if that little and tedious (doing it is tedious) information is expanded, even better. This is a long-term project; we do not have to finish everything quickly within a few weeks (8th Jan 2008 at 23:30).

ES9, also in favour of the policy, makes a final attempt to win votes in thread 1.14:

> What you all do not realise is that now you can create 20,000 articles with bots because there is no policy regulating their usage. This policy seeks to restrict large-scale creations, so that only those [articles] approved [by the community] will be created (9th Jan 2008 at 13:40).

As these last quotations illustrate, the negotiation behind the approval of the bot policy in the Spanish Wikipedia was marked by clear instances of rebellion. Editors against the policy showed a strong, often factionalist opposition on the grounds that having an official policy would encourage the proliferation of bot-generated articles. Some editors in favour of regulation believed that the aim of the policy was precisely to discourage massive bot creations. In the end, the divide between the two sides is not clear-cut. Regardless of their differences, both groups ultimately share the same aim: to improve the community.

In this debate, there are no peripheral members. The editors' date of registration and their narratives suggest that they are knowledgeable about Wikipedia. Since, as was stated in 2.2.1, only full members (registered users with more than 100 contributions) can vote in Wikipedia, it is not unusual to find that most discussants fell into this category. As will become apparent later in this section, this editor profile contrasts with the novice profile of some of the editors engaged in the discussion of the translation guidelines, where the formulation of policy is not at stake. For this reason, the insinuation by an editor in thread 1.10 that 'in this poll – more than in any other, you can see 'two sides': the veterans (old school) and the new Wikipedians' (ES10, 8th Jan 2008 at 17:09) is quickly dismissed by a senior editor endorsing the policy: 'Your phrase is very unfortunate. Many editors endorsing bot usage are veterans' (ES11, 9th Jan 2008 at 17:46).

Thus far, the negotiation process has proved unsuccessful. Editors on both sides have expressed irreconcilable, factionalist views on bot-generated articles and the need of regulation. Nevertheless, the situation is markedly different when what is being negotiated is not the existence of the policy itself but its content. In thread 1.21, entitled '*Sobre la política*' [Regarding the policy], ES12 challenges particular

aspects of the policy dealing with the requirements that potential bot runners have to fulfil to be granted permission to create articles. In their comment, ES12 draws attention to two points:

1. *Bot runners should be able to prove their knowledge of the policies, as well as their expertise in using bots. They should also be able to prove they are academically versed in the subject of their project.*

   Please, *Citizendium habemus*? Academic degrees? Excessive bureaucracy in my view. Can we remove that?

2. *The number of [bot-generated] articles could be greater provided that more editors get involved and that they are reputed senior members [of the community].*

   Prestige? How do we measure that? By the number of featured articles written [by the applicant]? By never having been blocked [in Wikipedia]? – 11th Jan 2008 at 15:00 (CET)

ES12's reference to Citizendium is not arbitrary. Citizendium was launched in 2007 by Wikipedia co-founder Larry Sanger, after leaving the encyclopaedia over disagreements with Jimmy Wales. Largely inactive today, Citizendium is also a user-driven, peer-produced encyclopaedia but, unlike Wikipedia, its editors are required to register an account using their real names and show their academic credentials ('CZ:Policies' 2020). Thus, the requirement to prove academic expertise goes against the spirit of Wikipedia. This was noted by O'Sullivan (2009, 88), who observes that the encyclopaedia's success resides precisely in the fact that 'contributors are judged by their track record of service to Wikipedia rather than by any formal or "real life" qualifications'.

Aware of the implications of sharing such information in Wikipedia, the main author of the draft, ES4, agrees to the proposed changes but delegates that task to other members of the community: 'As I said to another editor earlier, you are free to improve whatever you want' (11th Jan 2008 at 16:04). ES12 seizes the opportunity and changes the text (Figure 4-1), with neither ES4 nor other editors raising any objections. In the modified version shown or 'diff'[29] in Figure 4-1, whilst applicants are no longer expected to be reputed members, there must be no doubt as to their engagement with the community. Likewise, in the revised version

[29] In Wikipedia, 'diff' stands for 'different version'. Thus, it is possible to track changes undergone by a page only by selecting and comparing 'diffs' in the revision history.

applicants do not need to hold an academic degree, but they still should prove some knowledge of the subject.

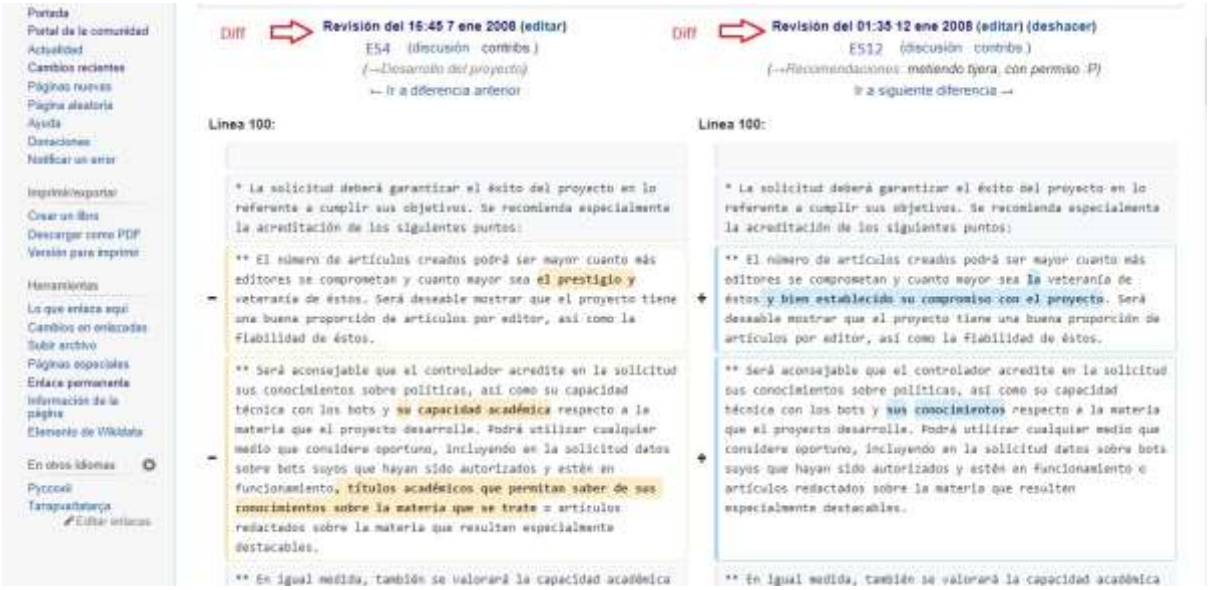


Figure 4-1. 'Diffs' retrieved from the revision history of the bot-creation policy page featuring changes to the text resulting from the agreement reached on the talk page.

Following the approval of the policy on 20th January 2008, there were only three requests to create articles using bots, one of which was rejected ('Wikipedia:Creación de artículos con bot/Solicitudes' 2019). The last successful application was on 7th February 2009. In addition, there were two attempts to modify the policy by voting, one in 2008, soon after the text was approved, and another one in 2015, neither of which was successful ('Wikipedia:Votaciones/2008/Sustitución de la política de botopedia' 2018; 'Wikipedia:Votaciones/2015/Modificación a la política de bots' 2019). Regulation seems to have fulfilled its purpose, preventing multiple creations despite fears by some that bots would have a negative impact on the community's practices. The last comment posted on the talk page, in thread 1.45, exemplifies not only distrust, but also an inability to disassociate bots from negative connotations. ES13, upset because they missed the poll, concludes: 'As usual, I have missed an important poll again, and all because the announcement board is not easy to find [in the village pump]. These *bots* may have won the battle today, but not the war' (31st Jan 2008 at 06:24).

In the second debate, which concentrates on the translation guidelines, ES14 (r. 2008)[30], a novice editor with fewer than 50 contributions in six years, initiated a thread on the talk page associated with the document *Ayuda:Cómo traducir un artículo* [Help:How to translate an article]. In their posting, entitled '*Uso de fuentes fiables en un artículo traducido*' [Use of reliable sources in a translated article], ES14 asks fellow practitioners whether it is true that 'as long as the source is reliable, it does not matter the language in which it is written' (19th Aug 2014 at 07:09). Their question is followed by a request: 'If so, could you specify that in *Ayuda: Cómo traducir un artículo*?' Six days later, ES15 (r. 2007), a Wikipedia administrator, replies that 'certainly, you can use references in other languages, although it is always desirable to find one in Spanish' (25th Aug 2014 at 14:05). Although an experienced member of the community answers the question relatively quickly, the request to add that information to the translation guidelines remains unattended at the time of writing this chapter.

This first interaction between the two editors epitomises the significance of learning in practice, but it also brings to the surface discrepancies as to how much of the knowledge is assumed. ES14, unversed in the topic, receives input from an experienced member of the community, who also gives them some advice. However, the same editor dismisses their petition, and no further action is taken. ES14's question and subsequent request show that incorporating references in other languages into the translated text is not something that less experienced editors can infer from the text. In brief, that knowledge is implicit. This implicitness is corroborated when ES15 answers the question and opts to keep the guidelines as they were. In doing so, ES15's actions raise an important question that Chapter 5 will endeavour to answer: how does that implicit knowledge transpire?

A second thread, created on 17th January 2018 by ES16 (r. 2010, 12 edits), is also illustrative of the explicit/implicit duality of competence. Again, an inexperienced editor asks other members of the Spanish Wikipedia community for further clarification. ES16's enquiry reads as follows:

[30] Due to the small number of participants in this discussion, editors are introduced directly in the text, along with their date of registration in brackets.

> From what I have seen, this article [page] covers the translation of articles from other language versions of Wikipedia as a process for the creation of articles in the Spanish Wikipedia. My doubt is: if the article already exists in Spanish, how do we proceed? Do we rewrite the article in Spanish using the translated text? Do we restrict ourselves to expanding the article in Spanish and leave the existing content just as is? (11:43 17th Jan 2018 (UTC))

On the same day, their questions are answered by an experienced editor, ES17 (r. 2016):

> Good question. It depends on the length, status and references that the article has. If you give me a concrete example, I can give you a more accurate answer. When it comes to expanding sections, I recommend translating the content in your user subpage first and then copy paste it (with the corresponding attribution: indicate where the information comes from, for example, "translated from the English Wikipedia" plus link). – (14:58 17 Jan 2018 (UTC))

These two postings reveal that the guidelines are unclear about what is the best course of action when the article that one wants to translate already exists in the Spanish Wikipedia. Once more, it is a beginner or peripheral member of the community who asks for guidance. ES16, despite being registered on Wikipedia since 2010, only made 12 contributions to the project in eight years. This lack of active engagement is what situates them in a peripheral position within their community and inbound trajectory. ES17, active only since 2016 but with many more contributions to the encyclopaedia, is well placed to answer the question. As with the 2014 thread, the response given by the experienced member demonstrates their knowledge of Wikipedia's policies, i.e. acknowledgement of the source. Although ES16 never asked about attribution, ES17 feels compelled to pass on the knowledge to them and ensure compliance with the standards.

This last subsection has shown that, on the surface, translation guidelines have not generated much debate within the Spanish Wikipedia community. The remarkably low levels of activity indicate that, at least in principle, the guidelines are not contested. Notwithstanding, there have been concomitant requests by inexperienced editors to clarify certain aspects of the text. Whilst fully-fledged members of the community have attended to such requests, no amendments were made to the guidelines. As far as the Spanish Wikipedia is concerned, the situation is different when it comes to local policies. As the bot debate revealed, Wikipedia editors tend to engage more eagerly with one another in internal policymaking processes where

there is something at stake. This is because, as was explained in 2.2.1 and 3.2, Wikipedia policies require certain consensus and the community must approve them ('Wikipedia:Policies and Guidelines' 2021). The next sections will show that, whilst participation is still relatively low, translation standards have engendered more debate in the other three Wikipedia language communities.

### 4.3.2 Negotiating templates and automation in the French Wikipedia

As noted in 4.2.3, the French Wikipedia lacks a specific policy regulating bot-generated content. Most translation-related threads initiated on the talk pages show, however, a high degree of engagement in matters pertaining to references and the use of automation. The data gathered from the talk pages linked to the documented standards analysed in previous sections reveal certain similarities between the French and Spanish Wikipedia communities. Much of the first part of the discussion revolves around the acknowledgement of the sources and the need to clarify specific aspects of the guidelines that lend themselves to more than one interpretation. The second part of the discussion focuses on automatic translation in general and the use of CX in particular. This subsection also seeks to ascertain the extent to which the editors' engagement leads to tangible amendments to the standards.

At the time of writing, there are ten discussion threads on the talk page ('Discussion aide:Traduction' 2020). Although the guideline was created on 5th November 2007, the first editor to post a comment on the talk page did so on 7th May 2015. Of the ten threads, only five discuss the document: four bring out possible ways of acknowledging the sources and another thread centres on automatic translation. In keeping with the criteria adopted in 4.3.1, the participants are listed in Table 4-4 below.

| Editor | Year of registration | Status within the community |
|---|---|---|
| FR1 | 2010 | Administrator |
| FR2 | 2012 | Experienced editor |
| FR3 | 2008 | Administrator |
| FR4 | 2015 | Experienced editor |
| FR5 | 2008 | Experienced editor |
| FR6 | 2007 | Experienced editor |
| FR7 | 2007 | Experienced editor |
| FR8 | 2004 | Experienced editor |
| FR9 | 2004 | Experienced editor |
| FR10 | 2008 | Administrator |
| FR11 | 2012 | Experienced editor |

Table 4-4. List of French Wikipedia editors that participated in the debate.

The debate about acknowledging the source began on 26th April 2016 at 14:07 (CEST[31]), when FR1 opened the thread entitled '*Nouvelle procédure concernant le crédit d'auteurs*' [New procedure concerning acknowledgement of the authors]. In their message, FR1 observes:

> Hello, according to the content translation tool FAQ section, it is no longer necessary to add the {{translation}} template to the article (or the talk page) if there is one link to the source in the comment box of the [edit summary] diff accompanying the translation. If that is the case, it will be necessary to update the page and spread the message as widely as possible.

The above comment is illustrative of how CX, an automated device, configured the practice of translation. When uploading the translation to Wikipedia, CX already provides a link to the original article, thus meeting the criteria set by the guidelines. This automatic process makes the use of the acknowledgement templates redundant.

[31] Unlike the Spanish Wikipedia, which observes Coordinated Universal Time (UTC), the French Wikipedia uses Central European Time (CET) and Central European Summer Time (CEST).

However, FR1's advice does not apply to translations performed without the device. In those cases, the options are less clear, and editors can acknowledge the source in at least three different ways ('Aide:Traduction' 2020). This lack of clarity prompted a lengthy discussion spanning almost four months, from 22nd May to 3rd September 2016. The thread, entitled '*Modèle obligatoire*' [Obligatory template], opens on 22nd May 2016 with FR2 requesting an explanation: 'I want someone to show me the obligation to use these two templates'. This request comes after FR2 observed that not all members of the community are following the recommendations: 'Why do these templates exist if nobody puts them?' (22nd May 2016 at 18:54)

FR3 attends to the request by referring to the previous thread: 'The templates for acknowledging authors are optional only when the translation is done with the automatic translation tool [CX]' (24th May 2016 at 04:46). Five weeks later, FR4 expresses their disagreement over how the explanations are phrased in the guidelines:

> I disagree with the text: 'It is obligatory to credit the authors using the two templates'. It seems to me that there are three solutions (link to the original article in the edit summary, use Translation/Reference, use Translated from [template]) and that it is up to the editor to choose to apply just one, two, or the three of them. As indicated on the *Aide: Crédit d'auteurs* [Help: Acknowledgement of the authors] talk page, it is sufficient to use only one of the three solutions (30th Jun 2016 at 23:21).

Unconvinced by the previous statements, FR2 addresses their colleagues and reinforces their initial request:

> [FR3] and [FR4], as you wish, but it just needs to be clear to beginners that the same is said on those help pages. Personally, I have been told to include the two templates; I did not know they were not obligatory (although it seems odd not to credit the authors with the templates, I have just seen that [FR3] says that it is okay).
>
> If that is the case, you should modify the help pages accordingly and stop telling newcomers that the templates are mandatory. I leave it to you (11th Jul 2016 at 11:38).

FR2's comment is revealing of their commitment to the French Wikipedia community. The editor not only reflects on one aspect of their practice – using the templates – but their senior status leads them to adopt a mentor role, showing concern for other less experienced editors. FR2 believes that the standards involving the use of translation templates are not explicit enough. Despite their

scepticism about the non-obligatory nature of acknowledging the source, FR2 trusts the feedback provided by their colleague and leaves the final decision in their hands. The imperative 'stop telling newcomers that the templates are mandatory' situates FR2 as a preoccupied editor who feels entitled to request changes to safeguard the successful inbound trajectory to membership of new practitioners.

Upon reading FR2's request, FR4 decides that it is prudent to wait for FR3's confirmation before proceeding with any changes: 'Before correcting this help page, it is better to wait for [FR3] to confirm whether or not there are three solutions' (11th Jul 2016 at 12:28). This exchange between FR2 and FR4 suggests that both editors are familiar with one another and have crossed paths with FR3 in previous debates, which would also explain the message 'I leave it to you'. FR3, validated by their peers' recognition, posts their answer two hours later, stating that: 'The link in the diff summary is the minimum legal requisite. As for the rest, these are *conventions* established by the community, probably out of habit' (11th Jul 2016 at 14:06). The author's use of italics on the word 'conventions' stresses the normative nature of practices. Admittedly, their repeated performance within the French Wikipedia community has resulted in their conventionalised usage.

Despite their initial claim, FR3 retracts their statement the following day, now holding that: 'I discovered that it is compulsory to use both templates. These templates are there as a favour [to the translator], but their obligatory nature (from where, by the way?) is not obvious to me, although I use them quite often' (12th Jul 2016 at 00:51). Although FR3 does not provide more details on how they reached that conclusion, their comment draws attention to the implicit knowledge of practice. FR3 is compliant despite not knowing who decided that using two templates was the best course of action. Following their remark, FR1 proposes updating the guidelines so as to leave no doubt about the obligation to add two acknowledgement templates to the translated article (12th Jul 2016, 01:01). As noted by O'Sullivan (2009, 108), the conundrum of trying to find which policy or guideline to follow is common in a bureaucratised system such as Wikipedia, where 'onsite conventions, policies and advice have accumulated over the years'. Wenger, McDermott, and Snyder (2002) refer to this phenomenon as documentism, a practice-related disorder that communities can overcome by discerning genuinely useful documents from those that are not, as seems to be happening in this debate.

The debate reached its peak on 3rd August 2016 when FR1 suggests giving priority to the option of including the {{Translation / Reference}} template in the 'References and notes' section of the translated article because it is more visible to the readers than the other two solutions. FR1 goes on to say that 'one thing that bothers me about crediting the [original] author in the [revision] history is that finding the oldest version [in the history] is a task for wiki-geeks' (3rd Aug at 21:44), hence not intuitive to peripheral members or newly recruited practitioners. As Wenger, McDermott, and Snyder (2002) observe, some practices, regardless of their efficiency, can pose challenges to outsiders and create boundaries for and among practitioners. As for including the template on the talk page of the translated article, FR1 observes that 'most readers do not check talk pages and revision histories (and certainly very few are interested in author credits)' (3rd Aug at 21:44).

In response, FR3, who had been hesitant about the number of translation templates that should be used, expressed their agreement with FR1's statement and concluded that having the template in the 'References and notes' section of the article is 'the best practice' (4th Aug 2016 at 10:22). As a result, the guidelines were updated (Figure 4-2), although not until almost two years later, on 25th February 2018, now explaining that crediting the source exclusively in the revision history is discouraged. The introduced changes – in bold – also highlighted that, despite not being compulsory, it is recommended to add the 'Translated from' template to the talk page. The anonymised editor FR Y wrote the message 'clarification regarding the last solution' in reference to the agreement reached on the talk page. This thread, summarised for the purposes of the analysis to include points relevant to the investigation, exemplifies the importance of finding proper ways of acknowledging someone's work in Wikipedia.

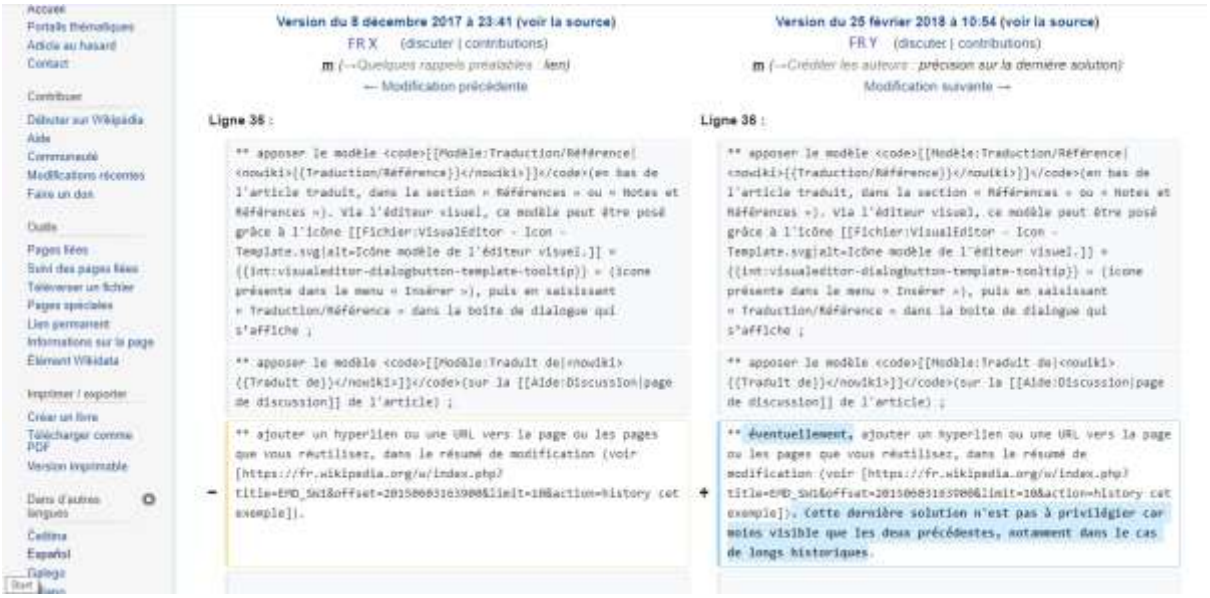

Figure 4-2. 'Diffs' retrieved from the revision history of the 'Aide: Traduction page'.

Aside from the more technical aspects of Wikipedia translation, another topic mentioned on the talk page concerns the use of automated devices. Once more, FR2 tries to negotiate the advice given to beginners in the translation guidelines, more specifically the following phrase: 'Out of respect for the readers, it is important never to perform an automatic translation. Any article created automatically will be deleted'. In their posting, FR2 criticises what they refer to as the 'demonization of machine translation':

> I have done all my translations using Google Translate and Linguee. I have never received feedback from anyone saying that that was not French (I am not saying my translation is perfect either). I think it is a shame to pull the rug out from under the people's feet by constantly demonising automatic translation tools. I would NEVER have translated anything without automatic translation.
>
> I think it should be rephrased and say something like:
>
> You can use automatic translation tools such as Google Translate or Linguee.[32] But be careful, automatic translation tools must remain aids, so you should never copy-paste unrevised automatic translations directly into Wikipedia (2nd Oct 2015 at 18:35).

[32] Although FR2 refers to Linguee as an automatic translation tool, it is in fact an online bilingual concordance.

The editor's choice of words is quite revealing. By criticising the 'demonization' of automatic translation, FR2 calls for a major integration of these materials into the community's practices. The fact that FR2 chooses to capitalise the adverb of frequency 'never' (*jamais* in French) to reflect on their performance – as something they would not have done – reinforces their disapproval. As became apparent in the previous thread on the obligation of using acknowledgement templates, FR2 is committed to changing the practices of their Wikipedia community. They do so by drawing attention to specific sections of the guidelines that, in their view, need revision. Although no one replied to FR2's comment on the talk page, the phrase was modified on 17th April 2017 at 16:17 by FR3. The recommendation to avoid automatic translation devices was not removed. Yet, the updated and current version presents a better picture of automation, one that has been disassociated from its previous negative connotations: 'Translation tools can *facilitate work*[33], as long as the generated article is revised and its content is thoroughly adapted [to Wikipedia]' ('Aide:Traduction' 2020)

Considering that one and a half years elapsed between the formal request and the actual change in the guidelines, it is not possible to conclude whether such modification was prompted by FR2's observation or a major acceptance of the role of automated devices in configuring practices. While that could well be the case, it is worth noting that Wikipedia's CX, despite being launched in 2014, did not become available for French Wikipedia editors until 2015. Therefore, when FR2 posted their comment, the device was still unknown to many editors. As noted in 4.1.6, French Wikipedia ranks second behind the Spanish Wikipedia in the number of articles created using CX. The quantitative data retrieved from the tool statistics indicate that CX has been successfully incorporated into the French Wikipedia. Since its inception in 2015, CX has been widely mobilised by French Wikipedia editors. This mobilisation could have contributed to automation being regarded more positively (disassociation) among senior members of the community, resulting in changes to the translation guidelines to reflect the new situation. Chapters 5 and 6 will seek to shed light on this issue with data obtained from the interviews.

[33] My emphasis.

These changes are also perceptible in the debate that transpires on the talk page attached to Wikipedia:Automatic translation ('Discussion Wikipédia:Traduction automatique' 2020). In 2010, five years before CX was launched, FR5 initiates a new thread entitled '*Essai?*' [Essays?]. In their comment, FR5 challenges the necessity of having a documented standard giving advice on how to use automated devices when, in their view, they should be avoided altogether:

> It seems to me that everyone agrees on not using automatic translation, especially since it stems from the most obvious common sense. In fact, articles created this way have been deleted immediately.[34] Is it [then] really necessary to engage in a decision-making process on giving recommendations in the case of something so trivial? (26th Feb 2010 at 15:15)

FR5 opens the debate the same day the *Traduction automatique* page is created under the title *Essai: Traduction automatique* [Essay: Automatic translation]. Their observation receives the support of another user, FR6, who adds that '[T]he problem with this text is that in a way it validates the practice of something that is discouraged' (26th Feb 2010 at 15:28). Their criticism prompts other experienced editors to participate and negotiate the relevance of having a page on automatic translation. Thus, FR7 replies:

> The problem, [FR6], is that (if no problems arise) we are heading toward the 22nd century. WP [Wikipedia] has the moral duty (though not the obligation) of anticipating the novel tools that will be available to future translators who have not necessarily been involved in English-to-French [Wikipedia] translations (26th Feb 2010 at 16:30).

As the above comment reveals, FR7 believes that the user-generated encyclopaedia should cater to future generations of translators. The last part of their response focuses on the status of those translators within the French Wikipedia community. According to FR7, the addressees of the *Traduction automatique* page need not be members of the community. Therefore, in their understanding of Wikipedia's mission, the translation standards set by the editors transcend the limits of their community of practice. In other words, the advice on how to handle automatic translation devices applies to both members and outsiders. FR7's message also

[34] The original French sentence is '*[ils] sont déjà passés par les armes sur le champ*'. It is an idiomatic expression meaning 'execution in the battlefield by a squad'. Whilst the idiomatic value is lost in translation, the intended meaning of 'making something disappear' was kept.

recognises that practices are dynamic and in constant flux, thus implying that fighting against their evolution is a futile endeavour.

The debate on the practicalities of automatic translation devices continues in a new thread entitled '*Ma façon de travailler*' [My way of working]. On this occasion, FR8 reflects on their own practice, arguing that 'automatic translation devices can be good tools, just like spelling and grammar checkers' (27th Feb 2010 at 05:32). They also hold that while programs such as Google Translate are like 'advisors whose opinions are not to be dismissed', the human editor 'should always remain the master of the game' (27th Feb 2010 at 05:32). FR8 also brings to the fore common errors associated with machine translation. Having translated many articles about medicine in Wikipedia, FR8 mentions cases where Google Translate is highly inefficient and critical judgment is required. For instance:

> 'Spinal disc herniation' is translated by Google as '*hernie discale épinière*'; the expression seems suspicious to me, and I realise that indeed the correct name in French is '*hernie discale*', an expression that I also find in my dictionary (27th Feb 2010 at 05:32).

In the same thread, FR9 concurs with the previous observations and concludes that:

1. The use of automatic translation alone is not practical: it should at least be accompanied by a customizable spelling and grammar checker that corrects with ease the expressions that must be revised while allowing the machine to acquire new words and expressions.
2. The importance of the subject matter (in your example, medical): this is where the translator's intelligence is essential because the automatic translation tool cannot recognise the semantic field. In this case, the translator herself must know at least the topic.
3. Automatic translation devices are tools, but not magic wands (as many people and IP addresses use them) (1st Mar 2010 at 12:15).

These comments illustrate what is expected of translators in Wikipedia. In their last sentence, FR9 criticises how both registered and anonymous (IP addresses) editors misuse the device in the French Wikipedia. They do so despite the community's best efforts to establish guidelines and policies on how to translate. Chapter 5 will delve into what may drive certain editors to ignore or defy conventions.

Earlier in this section, it was argued that the implementation of CX could have contributed to a more favourable view of automated devices in the French Wikipedia. The comment posted by FR7 above indicates the necessity of incorporating automated devices, especially as that could help future generations of

translators. Although the comments posted in 2010 by FR8 and FR9 show certain distrust of devices, a thread initiated seven years later on the same talk page presents a better picture. FR10 starts a new discussion thread entitled '*La situation a évolué, heureusement*' [The situation has improved, fortunately]. In their posting, FR10 notices that:

> The situation has improved: of course, there are always preposterous translations. Despite that, we still quite often find fairly correct translations whose length can reach a full sentence! Provided, of course, that behind a seemingly correct sentence there is not a huge misunderstanding... (21st Dec 2017 at 21:24).

This last comment reveals a shift in perception of automated devices among some members of the French Wikipedia community over the last lustrum. The impact of CX on translation practices in the community is attested by the creation of the dedicated page on 31st March 2015 ('Aide:Outil de traduction' 2020). On the ancillary talk page, a message posted by FR2 recognises the importance of having a help page devoted to CX ('Discussion aide:Outil de traduction' 2020). The editor validates their colleagues' work by praising them for having created a page that does not exist in other Wikipedia language communities, and that is likely to become 'an international reference for documentation' (23rd Apr 2015 at 18:04).

At the time of writing, there are eleven discussion threads, of which only one involved the negotiation of the guidelines. In a thread initiated on 20th April 2015 and entitled '*Renommage*' [Renaming], a few editors criticise the misleading use of the word '*Aide*' [Help] in the page title because, in their view, the aim of the page should be to describe the device and its functionalities rather than providing guidance on how to run CX. For instance, FR11 argues that 'the titles of help pages should be as accurate as possible to facilitate their consultation' (20th Apr 2015 at 13:01). On reading this post, FR2 attends to their colleague's remark, with the result of the negotiation being a new Wikipedia page: '[FR11], I have just created a page *Wikipedia: Outil de traduction* [Wikipedia: Automatic translation tool] to differentiate them' (21st Apr 2015 at 21:01).

The remainder of the discussion focuses on the device and its impact on configuring translation practices. Recurrent themes are software bugs found while running CX between 2015 and 2020, the unavailability of the device in specific languages when

it was first launched, and advice on how to credit the source(s) using acknowledgement templates (‘Discussion aide:Outil de traduction’ 2020).

As this section has shown, experienced French Wikipedia editors engage with one another to put forward changes to translation standards. The correct use of acknowledgement templates and the impact of automatic translation, including CX, were the foci of engagement in most discussion threads. Automation seems to be an integral constituent of the French Wikipedia community’s shared repertoire. The editors’ joint enterprise is underpinned by their efforts to regulate the mobilisation of automated devices – through postings and the creation of a help page – as well as by finding proper ways of acknowledging other editors’ work.

In essence, this is not different from the discussion that unfolded in the Spanish Wikipedia, motivated by the necessity of regulating bot-generated content. Both Wikipedia communities tackled the impact of content-creation devices on their translation practices and sought ways to restrict their deployment. In the case of the Spanish Wikipedia, the community approved a policy imposing limitations on bot-generated content. In the French Wikipedia, editors joined forces to create a policy on automatic translation and a help page explaining how to mobilise CX. The postings analysed, however, show that negotiation is not always straightforward, as some fully-fledged editors did not hesitate to rebel against specific aspects of the guidelines. The analysis also indicates that at times dissent was more amicable, innovation-oriented, and constructive, resulting in changes to the pages. As the following sections will show, there are certain parallelisms with the Dutch and Swedish Wikipedia communities.

### 4.3.3 Negotiating bots and sources in the Dutch Wikipedia

As mentioned in 4.2.3, as of May 2020, over 50% of the Dutch Wikipedia encyclopaedic articles have been created using bots. These large figures are due to the fact that the community has been more open to their mobilisation. In December 2011, there was an opinion poll to gauge the community’s preferences

('Wikipedia:Opinielokaal/Botimport' 2011).[35] Editors were asked to vote for either of the two following options:

a) All bot imports of new articles must be submitted to the community in advance via the WP: VBA[36] page.
b) WP: VBA should be changed from a request page to a notification page.

As is common with many Wikipedia decision-making processes, only registered editors with more than 100 contributions to the project were allowed to vote. All editors were given one week to cast their vote, from 11th to 18th December 2011. The first option was endorsed by 38 editors (47.5%) and opposed by 40 (50%), with two editors (2.5%) expressing a neutral opinion. The second option asked editors to indicate if each project involving bot creations should be surveyed by the community or simply reported without seeking approval. Only eight editors (20.5%) endorsed the need for approval, 18 editors (46.2%) declared that reporting creations to the community was sufficient, and 13 (33.3%) did not have a strong opinion about the subject. Consequently, the outcome was that, out of courtesy, editors are encouraged to report all bot creations to the community using the WP: VBA page. According to NL1 (r. 2009)[37], who closed the poll on 19th December 2011 at 00:57 (CET[38]), 'although the [Dutch Wikipedia] community is divided on a number of issues, it seems feasible to reach a compromise by having a notification page' ('Wikipedia:Opinielokaal/Botimport' 2011).

Such compromise, however, does not mean that reporting to the community is required. As stated in 4.2.4, reporting bot creations to the community is not obligatory but, rather, it is considered a gesture of goodwill. On further inquiry, a Dutch Wikipedia administrator contacted via email on 24th April 2020 confirmed that:

> We do not consider our help-pages part of our rules and regulations; therefore, this option is considered more the polite thing to do than a rule that needs to be enforced. I do think all of our bot-users do this, though.

[35] The link to the poll as well as the translation of the relevant excerpts into English were provided by a Dutch Wikipedia administrator, who was approached via email on 24th April 2020.
[36] WP: VBA stands for '*Verzoekpagina botimport nieuwe artikelen*' [Request page for the creation of new articles with bot].
[37] As only one Dutch Wikipedia editor is cited in this section, no table was added.
[38] Dutch Wikipedia observes Central European Time (+1).

Following the desire of 31 of the 39 editors who voted in the poll, the Dutch Wikipedia community opted to be flexible about bot-generated content. This decision contrasts with the one reached by the Spanish Wikipedia community where, as discussed in 4.3.1, a heated debate unfolded, with the result that bot-generated content must be reported, regulated and closely monitored. This remarkable difference illustrates how each Wikipedia language community has freedom to establish its standards. Thus, although bots are operated in many Wikipedia communities, who has access to them is determined by the existence or paucity of regulation. As has been observed throughout this chapter, dissenting opinions by engaged core members are common, but policy enforcement in Wikipedia is ultimately a democratic decision. The outcome is binding and local; it is part of the community's shared repertoire. Therefore, other Wikipedia language communities are not expected to follow suit.

Negotiation over further aspects of the translation standards continues on the talk page attached to the guidelines ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). At the time of writing, there are seven discussion threads spanning eight years, from 2007 to 2015 ('Overleg help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2015). The last message was posted on 3rd January 2015. Of the seven threads, five centre on citing sources, one revolves around copying lists from other Wikipedia communities, and another one focuses on the use of internal links in translation. As per the criteria followed in previous sections, the list of participants is in Table 4-5 below.

| Editor | Year of registration | Status within the community |
|---|---|---|
| NL2 | 2006 | Experienced editor |
| NL3 | 2003 | Experienced editor |
| NL4 | 2005 | Experienced editor |
| NL5 | 2004 | Experienced editor |
| NL6 | 2002 | Experienced editor |
| NL7 | 2003 | Experienced editor |
| NL8 | 2004 | Experienced editor |
| NL9 | 2006 | Experienced editor |
| NL10 | 2009 | Experienced editor |

Table 4-5. List of Dutch Wikipedia editors that participated in the debate.

In the first thread, entitled '*Lijsten overnemen van andere Wikipedia's*' [Copying lists from other Wikipedias], NL2 complains about a common error apparently made by novice members of the community:

> Many new users copy lists from other Wikipedias without further editing. This means that afterwards, you have to make changes like these [link]. It might be helpful to place a 'warning' about this on the page (2nd Sep 2007 at 19:11).

NL2 provides a 'diff' to substantiate their claim. The 'diff' shows a list of Golden Globe-awarded films imported from the English Wikipedia into the Dutch Wikipedia without translating the page titles into the target language. NL2's

complaint and subsequent request to warn beginners against bad practices are attended to by the editor who created the guidelines. NL3 takes on the recommendation of their colleague and replies: 'Thank you. Done' (2nd Sep 2007 at 23:04). A 'diff' retrieved from the revision history of the page (Figure 4-3) confirms that, following NL2's advice, the guidelines were modified on 2nd September 2007 to include the intended message in the community's shared repertoire.


Figure 4-3. 'Diffs' retrieved from the revision history of the Dutch Wikipedia translation guidelines.

In a second thread entitled '*Links*' [Links], NL4 asks their colleagues about interwikis, links that appear on the left side of each Wikipedia page that connect that particular page with other Wikipedia language communities where similar content is available. In their question, NL4 reveals their lack of understanding of how interwikis are added to articles and seeks assistance from their colleagues:

> How do translations end up in the left column of the original article? Can anyone provide a link between the pages for Peter Hinssen (en|nl) and the Factory ship (nl|en)? – 3rd Jan 2015 19:07 (CET)

The editor's request, despite not being related to the guidelines, is illustrative of how even experienced Wikipedia editors struggle with technical aspects that are not intuitive and require familiarity with wikisyntax and devices. As noted by O'Sullivan (2009, 107), 'learning to master the complicated vocabulary that Wikipedians use can be daunting for a new recruit'. Five years have passed since

NL4's posting, and no one has attended to the request. The Wikipedia entries mentioned in their comment have not been linked to the original English Wikipedia articles either. It remains unclear whether NL4 has learned how to do it and, if so, why they have not added the links. The unattended message, however, suggests that the theme has so far been overlooked by the community and that certain issues are unresolved. This low level of engagement contrasts sharply with the much higher levels of activity registered around the citation threads.

As shown in 4.3.1 and 4.3.2, acknowledging the sources is something that goes beyond the limits of a specific Wikipedia language community. Thus far, all three communities analysed in this chapter have expressed a keen interest in finding proper ways of acknowledging the source of a translation. In the Spanish Wikipedia, the main concern was whether the translated article should contain additional references in the target language. In the French Wikipedia, the discussion centred on the obligatory nature of using more than one template to acknowledge the source. In the Dutch Wikipedia, the five threads that are examined below share the same concerns, but they do so from a different angle. While templates are merely mentioned in passing, attention is placed on whether the original sources should be imported to the translated article.

The debate around citations can be divided into two parts. The first part comprises two threads posted in 2009, and the second one contains two follow-up threads written in 2011. A fifth thread, entitled '*Macro's tbv het vertalen van 'Citeer'-sjablonen*' [Macros for translating 'Cite'-templates], refers to updates to citation templates and differs from the other threads in that it is communicative in nature. Its author simply informs their fellow Dutch Wikipedia editors about updates to the code (27th Nov 2009 at 10:57). The first of the 2009 threads is entitled '*Bronvermelding: (bv.) Engelstalige Wikipedia*' [Citing sources: (for example) English Wikipedia]. In their message, NL5 asks colleagues to modify the translation guidelines to include information about Wikipedia's license agreements:

> I have come across several [Wikipedia] articles with the following source reference: 'The current or an earlier version of this article is (partly) translated from the English Wikipedia, which is subject to GFDL.'[39]
>
> Is this the correct sentence? Can it, with a short explanation, be added to the relevant section on this help page? – 2nd Jun 2009 at 19:02 (CEST)

As occurred with other threads, the editor's request goes unnoticed. As a result, sixteen days later, NL5 decides to take the matter into their own hands. On 18th June 2009, they introduce the GFDL reference to the guidelines, but their changes are reverted by NL6 on 11th November 2009 with the message 'no support' in the edit summary ('Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia' 2019). This reversion is elucidative since it sheds light on a case of failed negotiation. An experienced editor, NL5, tried to negotiate changes to the guidelines by initiating a thread on the talk page. Nevertheless, as time went by and no one replied, NL5 interpreted the lack of objecting voices as a free pass to modify the help page. Their action was reverted five months later by NL6, who authored the second 2009 thread entitled '*Neem de bronnen van de andere Wikipedia over*' [Copy the sources from the other Wikipedia]. In this thread, posted on 11th November, NL6 comes up with a detailed explanation of what prompted them to revert the passage on bibliographic sources, including the embedded GFDL reference:

> This passage has been the subject of discussion multiple times. Each time the tip in question was judged negatively by a large majority. I quote from three rounds of discussion:
>
> • Ri.: 'Copying sources which you have not consulted yourself should not be allowed. It's bad advice'.
>
> • Br.: 'I support that view'.
>
> • P.B: 'Wholeheartedly agree'.
>
> • B. D.: Citing a source that you have not consulted (one that you have copied from someone else) – to me, it seems unacceptable, parading another's achievements.
>
> […]

[39] GFDL stands for GNU Free Documentation License, a copyleft license designed by the Free Software Foundation (FSF) for the GNU project. Readers can copy, redistribute and modify a work as long as the derivate is published under the same license.

> One can conclude that this piece of advice does not have enough support. I will remove the passage in anticipation of an alternative formulation. – 11th Nov 2009 at 10:56 (CET)

As the above posting illustrates, NL6 compiles a list of archived discussions and quotations from other Dutch Wikipedia editors who expressed their opinion about the topic in the past. The first list, partially reproduced here, contains the views of ten editors, including that of the author of the thread. Aside from nuances of meaning and different styles, all editors within this first category agree on the fact that importing references from the source Wikipedia article into the translation is not a good practice. The tone employed by the discussants varies, with some editors overtly expressing their disapproval and classifying the practice as dishonest. The second list shows a pool of three experienced editors 'who argued differently'. These editors think that copying the original sources into the translation is a good practice, even when the translator has not had first-hand access to the references consulted by the original author(s).

Once more, the option endorsed by the majority prevails. GFDL stipulates that everyone can copy and redistribute any work covered by that license provided that the derivate be published under the same terms. Such terms raise ethical issues since they allow Wikipedia editors to import the original sources into the translation without infringing any legal provisions while encouraging the incorporation of citations they have not consulted. Therefore, NL6's rationale for reverting their colleague's addition to the guidelines is motivated by the ingrained factionalist convictions of a group of editors who regard the practice as unethical. In other words, NL6's actions are prompted by what they think is in the best interests of the Dutch Wikipedia community. Their reversion is not contested by NL5, who simply pastes the removed sections of the text into the talk page 'for illustrative purposes' (27th Nov 2009 at 14:36). Notwithstanding, NL5 still holds that 'copying sources that are available online should not pose problems, as they are easy to consult/check' (27th Nov 2009 at 16:08).

The negotiation continues when NL7 intervenes and tries to reconcile the two opposing views:

> Agreed; when you translate an article from another Wikipedia, I believe it is self-evident that you copy the sources used, *provided that* you checked them yourself (31st Jan 2010 at 20:26).

Despite NL7's mediation, no further actions were taken by either party. It was not until 24th September 2011 when another senior member of the community, NL8, initiated the first of the two 2011 threads in reaction to NL6's removal of the text. The first thread is entitled '*De verwijderde tip over het overnemen van bronnen van andere Wikipedia's en de beperkingen hiervan*' [The removal of the tip for copying sources from other Wikipedias and its limitations]. NL8 contests the deletion of the passage by their colleague, accusing them of acting unilaterally:

> In his argument, [NL6] makes it seem as if there were only two positions in the previous discussion:
>
> 1. Copy the sources from the other Wikipedia.
>
> 2. Do not copy the sources from the other Wikipedia.
>
> I have read through the previous discussions (from April 2007, April 2008, and Wikipedia:The_bar#Substantiation|nov 2009) and believe that there are in fact four different standpoints to acknowledge:
>
> A. Always copy the sources.
>
> B. Only copy the sources if you have checked them.
>
> C. Only copy the sources if you have consulted them.
>
> D. Never copy the sources.
>
> The original 2006/07 guidelines in fact correspond to standpoint (B). NL6 has actually provided a simplistic sketch of the situation. In reality, no one in the further discussion agreed with him, and he still did as he pleased. I believe this should be investigated more closely. – 24th Sep 2011 at 16:24 (CEST)

By raising these accusations, NL8 adopts a factionalist approach (Wenger, McDermott, and Snyder 2002) and calls for the previous recommendation to be restored. NL8's accusations against their colleague are quickly dismissed by NL9:

> Dear [NL8], it seems to me that you are putting words in [NL6's] mouth and misrepresenting their views. The only motivation for the edit I perceive (in the edit summary) is that there was no support for more. In addition, the situation is probably more nuanced. – 24th Sep 2011 at 16:35 (CEST)

Following NL9's response, NL8 tones down their previous statement and adopts a more conciliatory approach:

> Just to be clear: it is not my intention to scold NL 6. I am just assuming they acted sincerely according to their own convictions. They announced their proposal properly, and when no one protested, they made the change.
>
> Yet, I am still amazed by the result of this process, and I understand now how it has come about. I would appreciate if we could have this discussion again. Actually, I would like to propose putting the removed text back in place…unless there are better proposals. – 24th Sep at 17:13 (CEST)

In the second thread, entitled '*Neem onder bepaalde voorwaarden de bronnen van de andere Wikipedia over*' [Under certain circumstances, copy sources from the other Wikipedia], NL8 puts forward their proposal:

> On 11th November 2009, as seen here, the following text was removed. In response to previous points of discussion, I want to propose that the following text is once again included in the article:
>
> *Copy the sources from the other Wikipedia… Be careful: always check whether used sources exist, whether they can be accessed (for online sources) and whether they correspond to the text. Reference may be to an offline source, but you can check if it is available on the internet, for instance, on Google Scholar or Google Books. If you really cannot verify the sources, you can, depending on the text, consider removing this part of the text or to include it without source references. Of course, you can also consider visiting a library or bookstore with a good range of international reading matter.* –24th Sep 2011 at 17:18 (CEST)

Nonetheless, their attempts at having the 'discussion again' and negotiating changes to the guidelines fail when only one editor expresses their view. On 28th September 2011, NL10 rejects their proposal, arguing that it is not feasible: 'You are presenting the exception as a rule. The main rule should be: only copy sources that you have consulted yourself'. This last statement puts an end to a short-lived 2011 discussion that resulted in no changes to the guidelines. NL8's unsuccessful attempt and the lack of community response are an indication of marginality. Wenger, McDermott, and Snyder (2002) refer to marginality as a community disorder and observe that members whose opinions are ignored or silenced tend to lose interest over time as they experience 'the lack of effectiveness in making a difference'. The absence of debate in over nine years exemplifies this disengagement and may be an indication of passive conformity.

At the time of writing, the 'Referencing' section of the Dutch Wikipedia translation guidelines still discourages the practice of importing references without previous consultation:

> On the other hand, when translating, you do not have to cite sources that you have not consulted yourself, even when they are referenced in the original article. This is because you cannot be sure the sources support the argument presented.

As this section of the analysis has shown, not all editors in the Dutch Wikipedia agree on the standards of performance. This disagreement is noticed in both bot-generated content and acknowledgement of the sources. In the two cases, the view expressed by the majority was integrated into the community's shared repertoire. Such a repertoire differentiates one Wikipedia community from another. In the case of the Dutch Wikipedia, the brief, albeit heated, negotiation process that occurred between 2009 and 2011 on the translation guidelines talk page resulted in a series of changes to the standards. Consequently, the Dutch Wikipedia guidelines differ from the Spanish and French ones in that, at least as expressed in the text, the practice of copying non-consulted references is frowned upon. Having analysed the negotiation process in three Wikipedia communities, the next section moves on to investigate how standards are negotiated in the Swedish Wikipedia.

### 4.3.4 Negotiating bots and style in the Swedish Wikipedia

As noted in 4.2.3, the Swedish Wikipedia does not have a policy regulating bot-generated content. This information was confirmed by a senior administrator of the community, who was approached via email. In their answer, the Swedish Wikipedia administrator observed:

> We do not have a specific policy for bot-created articles on svwiki.[40] We have had one major project for creating articles by bot, and it was handled on a mostly ad hoc-basis until, unfortunately, it got very much out of hand, and the creator abandoned svwiki. There have been some bot discussions before and after this, but they are probably not easy to find.

The administrator also provided a link to the bot owner's user page, who collaborated in Wikipedia under the nickname Lsj. The editor's history of contributions confirms that they have been inactive since 15th November 2017, the

[40] Svwiki stands for Swedish Wikipedia, SV being the ISO code for *Svenska* (Swedish).

date of their last 'edit'. In their user page, the editor left the message '*Är inte längre aktiv på svenska Wikipedia*' [(I) am no longer active in the Swedish Wikipedia] ('Användare:Lsj' 2017). As explained in 1.1.2, Sverker Johansson (Lsj) is known in Wikipedia for being the developer and owner of Lsjbot, a cross-wiki bot that was first launched on 31st January 2012 to substantially increase the number of encyclopaedic articles in the Swedish Wikipedia. The bot made over 17 million contributions to the Swedish Wikipedia, creating more than 80% of the community's 3.7 million articles ('Lsjbot' 2021). Its last contribution was made on 13th November 2016 at 23:54 (CET).

As highlighted by the administrator in their comment, this automated device caused some controversy in the Swedish Wikipedia. As a result, the bot was ceased in 2016, and its developer left the community one year later. Although more insight into the controversy will be provided in 6.2, a comment posted on 1st May 2020 by another Swedish Wikipedia administrator on Lsj's talk page reveals that neither the editor nor the bot were banned from the community ('Användardiskussion:Lsj' 2020):

> Hi, Lsj! Since you announced that you are no longer active in the Swedish-language Wikipedia and you have not been there for three years and have not used Lsjbot either, I have revoked the bot flag according to the guidelines for robot users. In good faith, SV1, 1st May 2020 at 19:42 (CEST)

The administrator revoked the flag following the criteria set by the bot policy, which stipulates that bots that have been inactive for an extended period will lose their license ('flag') to operate in the Wikipedia community where permission was granted. The above posting proves that the automated device that created millions of articles was still allowed to operate until May 2020. Still, for some reason, the developer chose not to do so. As is clear from the administrator's message, the flag was revoked for prolonged inactivity. While Chapter 6 will shed light on this issue, at this stage, it is apparent that the lack of policies regulating bot creations in the Swedish Wikipedia could have contributed to the situation being out of control.

Turning now to the essay, the data show little discussion since the document was first published on 4th January 2007. At the time of writing, there are five inactive discussion threads on the essay's talk page. These threads were preceded by a series of comments posted on the top of the page. Of the five threads, one was initiated in

2009, three unfolded in 2011, and another one was posted in 2018, after a seven-year gap. Since the Swedish Wikipedia translation essay focuses almost exclusively on grammar and style, four of the five threads tackle stylistic issues. The only thread that does not discuss the contents of the page is the one posted in 2018, entitled '*En essä*' [An essay]. In this case, an editor observes that a few colleagues referred to the document in a debate that took place sometime in 2018 on the *Bybrunnen*, the Swedish Wikipedia village pump (16th Nov 2018 at 08:12). However, they do not provide any more information as to what was discussed. The list of discussants is found in Table 4-6 below:

| Editor | Year of registration | Status within the community |
|---|---|---|
| SV2 | 2004 | Experienced editor |
| SV3 | 2003 | Experienced editor |
| SV4 | 2006 | Experienced editor |
| SV5 | 2006 | Experienced editor |
| SV6 | 2009 | Administrator |

Table 4-6. List of Swedish Wikipedia editors that participated in the debate.

The first comments posted between May and August 2007 are in response to the creation of the essay itself. On 17th May, SV2 posts a message to congratulate their colleagues for the work they have done with the page: 'Clear and well-written instruction'. Two other editors join in the congratulations, with one of them requesting minor changes to the document. In their message, SV3 suggests changing the phrase 'do not translate word for word':

> I wonder if it is possible to rewrite 'do not translate word for word' to something that is not expressed with a 'not'. 'Think whole phrases, not individual words' or something? (3rd Aug 2007 at 21:27)

SV4, the editor who created the page, welcomes their colleague's advice but argues that the original phrase is intended to get the message through, considering that the addressees are beginners (3rd Aug 2007 at 21:27). The negotiation ends with no changes introduced to the essay. Two years later, SV4 initiates the first thread, entitled '*Tonerna strax ovanför A*' [The tones just above A]. In barely one sentence,

SV4 informs other members of the community about the changes they introduced to the section of the documented standard dealing with the translation of musical tones. They justify their action as follows: 'I cut that section down sharply, most of it is a repetition of what is elsewhere' (25th Nov 2009 at 09:31). No one objected to the modification, and the debate went cold until 2011.

In December 2011, SV5 commented under the title '*Jag lade till litet*' [I added a little]. Self-identified as a mathematician, SV5 reflects on their practice of writing articles about their field of expertise in Wikipedia. Drawing on their experience, SV5 modifies the advice given in the essay to include 'that it can sometimes make sense to use the revision history to get a better text to translate in a place where something went wrong' (15th Dec 2011 at 18:41). Thus, the editor's addition encourages novice translators to undertake further research and revise past versions of an article if any aspect of the ST is obscure.

The second 2011 thread, entitled '*Klockslag*' [Time], is out of all the threads the one that registered the highest levels of engagement. It also began in December 2011, when SV6 posted a message rebelling against the recommendations given by the essay on how to translate hours into Swedish:

> *The colon as a delimiter between hours and minutes is also the format stipulated by ISO 8601 (a standard for time and date values). According to the Language Council, this should not be done in Swedish.*
>
> Do the recommendations of the Language Council really outweigh an international standard? – 16th Dec 2011 at 10:16

The translation essay draws primarily on the advice provided by *Språkrådet*, the Swedish Language Council. The institution promotes the advancement and prescriptive use of the Swedish language, thus fulfilling a similar function to that of the *Real Academia Española* and the *Académie Française* for Spanish and French, respectively. In their message, SV6 overtly challenges the authority of the Language Council regarding time notations on the grounds that it goes against the recommendations 'stipulated by ISO 8601'. In doing so, SV6 advocates a standardised international approach. Their criticism of the guidelines prompts the original author, SV4, to argue in defence of the recommendations:

> The Language Council, of course, has nothing to say about the standard itself, in the same way that the standard has nothing to say about writing rules in Swedish. The idea of the ISO standard is not that all people on earth should start writing time and date in this way. –16th Dec 2011 at 12:52

Far from being convinced by the above comment, SV6 replies:

> I would nevertheless not recommend anyone to follow the recommendations of the Language Council in this case […]. I consider that a standard outweighs recommendations. – 17th Dec 2011 at 12:52

SV4's response one hour later shows that, whilst they acknowledge the practicalities of using ISO, they are still against introducing any changes to the document: 'It might be better for the world if everyone used ISO, but it is not Svwp [Swedish Wikipedia] that should go for it. I think so' (17th Dec 2011 at 13:44).

On reading the previous statement, SV6 now argues that there is not one but two rules:

> There is not one norm; there are two. The one advocated by the Language Council (with a dot) and the one that the SIS [Swedish Institute for Standards] advocates (with a colon). I have previously had the impression that the former is rarely used, but when I surf around a bit, I see that it seems to be evenly spread. Sweden's radio writes with a colon too. – 17th Dec 2011 at 18:53 (CET)

After a series of back and forth messages, SV4 opts to dismiss the claims made by their colleague and refuses to incorporate ISO regulations into the essay:

> I imagine that over time, the use of the colon will become the preferred norm but, so far, the dot is what is prescribed by the Swedish writing rules, the Government's writing rules, the language teachers of the Swedish Academy, and all other style guides I have come into contact with. – 20th Dec 2011 at 12:14 (CET)

SV4 bases the decision on custom. Their actions suggest that they feel entitled to certain ownership, perhaps because, as stated in 2.2.1, essays lack communal consensus and tend to reflect the opinion of a small number of editors. In their prescriptive view, the Swedish Wikipedia should abide by the rules of the Language Council. In another thread, entitled '*Semikolon*' [Semicolon], SV2, who previously praised SV4's work on the essay, asks for clarification on the differences in usage between English and Swedish: 'From the text: "Semicolon is not used in Swedish

at all in the same way as in English...” How then?’ (17th Dec 2011 at 00:41). This time, SV4 is more receptive and further elaborates on that point, providing more context to their original explanation in the essay. As Figure 4-4 shows, in the second ‘diff’ SV4 indicates that they are ‘*hyfsning*’ [trimming].


Figure 4-4. Diffs showing the changes introduced to the Swedish Wikipedia essay.

## 4.4 Chapter conclusion

The analysis of the data retrieved from the documented standards reveals that the four Wikipedia language communities share a series of principles and values. These are the acknowledgement of the sources, compliance with the encyclopaedia’s license terms, and the use of correct grammar and spelling. Differences arise, however, when it comes to the criteria for selecting an article for translation and whether or not one should import the original bibliographic references to the target article without having consulted them.

As the two stages of the analysis have shown, communities also differ sharply in how they have incorporated devices into their practices. While the Spanish Wikipedia has a solid policy regulating bot-generated content, communities such as the Dutch and Swedish Wikipedia lack standards in that regard. Regarding other automated devices, the Spanish and French Wikipedia communities have mobilised

CX widely, whereas its deployment in the Dutch and Swedish Wikipedia communities has been more modest.

The analysis of the second dataset has shown that negotiation of the standards can be an arduous undertaking. Although most documented standards are aimed at peripheral members of the community, amendments to their provisions are usually negotiated by experienced members. The data gathered from the four Wikipedia communities demonstrate that various engaged individuals take the matter seriously. For instance, instances of rebellion were found in and across the four communities. In the Spanish Wikipedia, experienced editors had clashing views on bot-generated content, with some editors exhibiting instances of factionalism. In the French Wikipedia, editors were divided on the obligation or optionality of using more than one acknowledgement template and on the advantages and downsides to mobilising automated devices. In the Dutch Wikipedia, the main cause of disagreement among editors was the ethical issues related to importing unrevised references of the source article into the translation. In the Swedish Wikipedia, there were discrepancies between editors on the advice given by the essay regarding time notations.

Despite occasional pockets of resistance, significant changes to the standards were usually informed by the opinion held by the majority. This fact became clear in most cases, with the notable exception of the Swedish Wikipedia, where the final decision rested with the principal author of the essay, who rejected the changes proposed by two of their colleagues. The data also showed that modifications to the guidelines is a slow process, where only a handful of editors intervene at specific points in time.

Having ascertained the extent to which translation standards are regulated and negotiated, the next chapter will investigate whether, and if so, how, such standards inform the practices of 16 experienced Wikipedia translators.

# Chapter 5. Incorporation of the standards into practice

Following the analysis of documented Wikipedia translation standards in the previous chapter, the aim of Chapter 5 is to determine whether, and if so, how, these are applied in practice. To this end, the next sections will tackle the issues raised by the third research question (RQ3): 'To what extent have experienced Wikipedia translators incorporated the standards set by their language communities into their practices?' In so doing, this chapter seeks to assess the significance of translation standards (policies, guidelines, and essays) for 16 Wikipedia translators of the four communities.

Emphasis will be placed on both implicit and explicit features of practice and how they are acquired through experience, beginning with the challenges the participants faced when they were newcomers. Such features were discussed in Chapter 2, and examined, albeit superficially, in 4.3, which centred on the negotiation processes transpiring on the talk pages. Although normativity and negotiation in practice were addressed in the previous chapter, their relevance and currency for Wikipedia translators mean that such concepts will be recurrent in the thematic analysis of translation-as-performance.

This chapter is divided into five parts. The first section builds on the participants' background information. Section 5.2 pays attention to what prompted them to become Wikipedia translators. Section 5.3 of the analysis delves into the participants' experience as translators, including the challenges they encountered in their inbound trajectory to membership. More concerned with documented standards of practice, section 5.4 shifts the focus to the policies, guidelines and procedures that participants prioritise when they translate. The fifth and last section moves on to examine the participants' knowledge and incorporation of the translation guidelines established by their communities. Since local Wikipedia standards regulate who has access to certain materials, this chapter paves the way for the analysis of the mobilisation of automated devices in translation, which will be the focus of Chapter 6. Drawing on the same dataset, Chapter 6 – and in

particular 6.3 – will also deal with the evolution of translation practices in Wikipedia.

## 5.1 The participants

As explained in 2.3, Wikipedia is a constellation of communities of practice, each of which is characterised by a series of common (overarching) but also distinctive standards that make up the shared repertoire. Chapter 4 examined the overarching values regarding translation and then moved on to analyse those unique to specific language communities. For instance, familiarity with the subject was desirable in the Dutch Wikipedia when selecting an article for translation. At the same time, other Wikipedia communities such as the French and the Spanish put a spotlight on the importance of choosing featured articles. Section 4.3 revealed that policymaking in Wikipedia often involves an arduous negotiation process, with some elements of the guidelines triggering lengthy debates that last weeks if not months. This fact became apparent in the negotiation of the bot-creation policy in the Spanish Wikipedia, the use of translation templates in the French Wikipedia, the previous consultation of references in the Dutch Wikipedia, and the preference for specific stylistic rules over others in the Swedish Wikipedia.

Despite these revealing findings, a new methodology and dataset are necessary to explain the importance that senior Wikipedia translators confer on the standards and how they have incorporated them into their practices over time. To shed light on this matter, 16 experienced translators – four per community – were interviewed between July and August 2020. In what follows, their experiences are examined, adopting a thematic approach consistent with the methodology outlined in 3.4. Because themes go beyond the boundaries of specific communities – as opposed to the local nature of negotiated standards – the structure adopted here differs from the one followed in the previous chapter, where the results were presented by language community. In keeping with the criteria introduced in 3.4.1 aimed at safeguarding the privacy of the participants, all of them were given fictitious, gender-neutral names.

### 5.1.1 Participants' background

The first thematic block in the interview guide (see Appendix III) was intended to elicit some background information on the participants' translation experience in Wikipedia. First, participants were asked about their date of registration in Wikipedia and their motives to engage in translation. Second, participants were encouraged to share details about the type of encyclopaedic entries they had translated and the criteria they followed when deciding which ones they wanted to translate. Finally, participants were asked to describe their experience as translators in Wikipedia and comment on challenges they had faced in their practice and how they overcame them. This section will expand on the first point and the information the participants provided in a separate questionnaire (see Appendix IV) before the interviews. Sections 5.2 and 5.4 will address the participants' motivation to engage in translation and the Wikipedia policies they usually prioritise.

As explained in 3.4.1, prior to the interview, participants were sent a brief questionnaire via email aimed at gathering information such as a) age group, b) education, c) knowledge of languages, and d) previous experience in translation. Their date of registration in Wikipedia was asked during the interview. Although most participants remembered when they created their Wikipedia account, some were only able to provide an approximate date. To ensure accuracy, the information they shared was later verified via their accounts' registration log. The collected data are included in Table 5-1 below. The participants are divided by Wikipedia language community. In line with the criteria adopted in Chapter 4, the bracketed ISO codes ES, FR, NL and SV are used.

| Participants' background | | | | | |
|---|---|---|---|---|---|
| **Name** | **Registration date** | **Age group** | **Education** | **Level of English** | **Translation experience** |
| Alex (ES) | 2007 | >50 | Architecture | EN-4 | Yes |
| Ariel (ES) | 2015 | 31-40 | Chemical Engineering | EN-3 | No |
| Cris (ES) | 2009 | 18-30 | Student (BSc) | EN-3 | No |
| Pau (ES) | 2009 | 41-50 | Civil Engineering | EN-3 | No |
| Ange (FR) | 2004 | 31-40 | Environmental Science | EN-3 | No |
| Dominique (FR) | 2004 | 31-40 | Computer Science | EN-3 to EN-5 | Yes |
| Maxime (FR) | 2008 | 31-40 | Mechanics/Digital | EN-4 | No |
| Sam (FR) | 2017 | 18-30 | Oceanography (PhD student) | EN-3 | Yes |
| Guus (NL) | 2002 | >50 | Physics | EN-3 | No |
| Jos (NL) | 2005 | 31-40 | Engineering | EN-4 | No |
| Leslie (NL) | 2012 | 41-50 | Law | EN-3 | No |
| Nik (NL) | 2005 | >50 | Economics | EN-3 | No |
| Alva (SV) | 2009 | 41-50 | Engineering | EN-5 | Yes |
| Charlie (SV) | 2012 | >50 | Education | EN-3 | Yes |
| Kim (SV) | 2006 | >50 | Geography | EN-5 | Yes |
| Robin (SV) | 2008 | 18-30 | Literature (MA) | EN-5 | No |

Table 5-1. Data collected on the participants' background.

As Table 5-1 illustrates, five participants responded that they were over 50, another five were between 31 and 40, and three answered that they were between 41 and 50. The remaining three fell under the category of 18 to 30 years old. In terms of education, 13 participants had a scientific background, with four of them identifying as engineers. The remaining three participants consisted of a high school teacher, a lawyer, and a literature graduate. When asked to rate their knowledge of English using the Wikipedia Babel scale (see 3.4.1), all 16 participants reported their level to be above EN-3 (advanced). In addition, three of the four Swedish Wikipedia translators rated their proficiency to be much higher, at the professional EN-5 level.

The last section of the questionnaire also required respondents to indicate whether they had any previous experience as translators before joining Wikipedia. Only six participants (37.5%) stated that they had prior translation skills. Of these, three – Alex, Sam and Kim – had worked as professional translators, while the other three – Dominique, Alva and Charlie – responded that they had been involved in volunteer translation activities. When divided per community, three participants – Alva, Charlie and Kim – were from the Swedish Wikipedia, two – Dominique and Sam – were French Wikipedia editors, and the remaining participant, Alex, was from the Spanish Wikipedia. None of the Dutch Wikipedia participants had translated before. Despite the diversity in the participants' translation background, all of them can be regarded as senior editors of their Wikipedia communities. The following section examines the participants' motivations to become mutually engaged in the practice of Wikipedia translation.

## 5.2 Becoming translators in Wikipedia

The previous section showed that most participants in the study had a scientific background, and six were acquainted with translation. Nonetheless, as stated in 4.3.1, qualifications are neither required nor expected in a collaborative project such as Wikipedia. If anything, anonymity is perhaps one of the encyclopaedia's most salient features (O'Sullivan 2009, 88). Although anonymity is consequential in Wikipedia, registered editors are still part of a community and, as such, they are under a regime of mutual accountability.

The next paragraphs bring to the fore how participants were recruited by their communities, focusing on Wenger's (1998) tenet of joint enterprise. In doing so, attention is laid on those aspects that motivated the interviewees to become translators in the online encyclopaedia. Table 5-2 below shows broad themes that emerged after coding the data.

| Motivation to become a translator in Wikipedia | |
|---|---|
| **Themes** | **Number of participants** |
| Personal interest in a topic | 13 |
| Sharing free knowledge and improving articles | 9 |
| Translating is more manageable than editing | 4 |
| Improving language skills | 3 |

Table 5-2. The participants' main reasons to become translators in Wikipedia.

As discussed in chapters 2 and 4, communities of practice are characterised by a shared repertoire and a joint enterprise (Wenger 1998; Wenger, McDermott, and Snyder 2002). Having a joint enterprise or common goal is crucial for a community to come into being and thrive (Wenger 1998). Interest and aims as driving forces are frequent themes in practice theory, regardless of whether one subscribes to Wenger's concept of community (Nicolini 2012; Shove, Pantzar, and Watson 2012). In an open collaboration project such as Wikipedia, interest in a given task or topic plays an essential role (O'Sullivan 2009).

Despite being accountable to one another, Wikipedia editors are volunteers driven by their interests. A recurrent theme in the interviews was a sense amongst participants that having a professional or personal interest in a specific area is necessary to become a translator in Wikipedia. For instance, Alex (ES), who identifies as a 'transcreator' rather than as a translator, said that they began translating articles about their profession to fill that knowledge gap in Wikipedia. Likewise, Ariel (ES) and Cris (ES) translated articles on engineering and astronautics, respectively. Dominique (FR), a computer scientist and free software advocate, began by translating 'some works from [Richard] Stallman[41] and the GNU project'. Along the same lines, doctoral student Sam (FR) translated articles about their research domains, biology and oceanography. High school teacher

[41] Richard Stallman is an American free software movement activist.

Charlie (SV) first dabbled in translation when they 'had the idea of using Wikipedia at school for the students'.

Regarding personal interest, Pau (ES) and Guus (NL) noted they had translated entries about photography and painting. Kim (SV) declared that what prompted them to become a translator was their passion for languages. Ange (FR) joined the translation community when they were invited to participate in the 100 Wiki days project, which consisted of creating at least one Wikipedia article per day for over three months. Since then, they have taken an interest in translating articles on Celtic nations in general and notable Irishwomen in particular to reduce the gender gap in Wikipedia. Guus (NL) also pursues the same cause-driven goal: 'I'm involved in the gender gap group, so the last few months I've been translating on female painters. Articles about female painters or other official artists'.

Other participants declared having multiple interests. Maxime (FR), for whom translation is ancillary to their editing activity, said:

> Well, I have three different fields I'm translating about. Mainly trains from North America; they have a lot of things already existing, so it's easier for me to just translate, and I also try to do stuff about... Well, famous people, mostly women from North America or strange topics you can find that are on English Wikipedia sometimes or Spanish Wikipedia. So… Yeah, I translate things because I like the topic.

Another interviewee, Alva (SV), when asked about what motivated them to become a translator, answered: 'I've translated quite varied articles. I checked what I've done with that program [CX]: people, a company, an electrical device, a statistics thing, and a sports club'. Nik (NL), whose first translations were also motivated by preference, observed: 'I came across some topics and said, "Okay, I want to translate that from English to Dutch because I like the article"'. Pau (ES), primarily interested in photography, expressed a similar view: 'I write about mostly everything, and I use all the articles in all languages'.

Besides having an interest in one or more topics, nine participants expressed a desire for improving articles and making information available to others. These findings coincide with the results published by previous studies on the motivations of volunteer translators (O'Brien and Schäler 2010; McDonough Dolmaya 2012; Olohan 2014; Cámara de la Fuente 2015). All of them reported that the vast

majority of volunteer translators are driven by their interest in sharing knowledge and supporting the aims of their organisation. From a praxeological lens, it can be argued that the translators' joint enterprise is aligned with one of Wikipedia's central tenets: ensuring that knowledge is accessible in as many languages as possible. When asked about their inspiration, Pau (ES) explicitly mentioned sharing Wikipedia's mission:

> My motivation is the purpose of Wikipedia, of creating a free database of knowledge for everybody and make it accessible for everybody; and the way, in my opinion, the easiest way, to approach it or to contribute to this target is translating.

Sharing free knowledge is appealing to some and can lure editors into improving articles. This fact became apparent in the answers provided by Alex (ES) and Kim (SV):

> So, one day I started editing articles that I found were incomplete or with wrong things or missing... with typos, or whatever. It came all of a sudden. I became a Wikipedia editor almost unwillingly. – Alex

> First, you discover one thing to do, and then you discover another thing to do, and then you discover a third thing to do. It's like a snowballing experience. – Kim

The 'snowballing experience' that Kim describes is in tune with the answers given by six participants. When asked the same question, participants indicated that sometimes they felt the urge to fill a knowledge gap and improve existing articles. As a wiki-based platform, the encyclopaedia uses internal links to connect different articles and other pages. When an article does not exist in a given Wikipedia community, the hyperlink appears in red, encouraging readers to become involved and create the content. This strategy worked for Sam (FR) and Alva (SV):

> Usually, I'm interested in translating English pages that don't exist in French and that are in my domain of research, more or less. So, it's what I do most of the time when I'm on Wikipedia. – Sam

> Translation... Well, if I saw missing articles, I would translate them. – Alva

As shown in Table 5-2, some editors chose to become translators for practical reasons. For example, four interviewees considered that translating articles in Wikipedia was faster and more straightforward than creating encyclopaedic content from scratch. Pau (ES) illustrates this point in their answer:

> And I also realised it is quicker to translate something, I mean, let's not talk about the quality of the translation, but it's better to take a Finnish text, or an English, German, French, whatever, and translate it into Spanish than starting something from scratch. I have created some articles also from scratch, but it was very special things.

Furthermore, three participants declared having a vested interest in improving their language skills. However, this interest was not what prompted them to become translators in the first place. Instead, participants viewed translation as a way to improve their language and writing skills. Ariel's (ES) comment below elucidates this learn-by-doing (Wenger 1998) approach:

> I wanted to get a better handle on... Because, well, in my field of work, I have to use several languages at the same time, so it was kind of practice to handle issues with diction, grammar... Because there are things that, well, one studies; but with languages, you enrich yourself through reading and using them. – Ariel (ES)

Despite coming from different backgrounds, the participants' answers suggest that most of them engaged in the practice of translation for similar reasons. Acknowledging that communities of practice are far from being homogeneous units, i.e., no two members are alike (Wenger 1998), the holistic concept of joint enterprise serves, however, an essential purpose: it allows individuals to engage with one another and negotiate meaning. In this case, despite the fact that not all participants share the same interests, their efforts and goals converge. With this in mind, the following section examines the participants' inbound trajectory, focusing on the challenges they had to overcome when faced with their community's shared repertoire.

## 5.3 Becoming familiar with the shared repertoire

As explained throughout various sections of Chapter 2, once recruited by their communities, newcomers engage in a learning process through which they become full-fledged performers of their practice. Learning by doing takes on a central role in practice (Wenger 1998; Shove, Pantzar, and Watson 2012; Olohan 2021). Therefore, it is not uncommon for new members to be in a peripheral position within their practice and learn through their errors, often benefiting from the input received by more experienced, senior practitioners (Wenger 1998; Warde 2005). In what follows, the emphasis is placed on the participants' account of their

socialisation into the practice. The analysis includes, but is not limited to, the interviewees' interaction with materials that are part of their community's shared repertoire and the difficulties they encountered when they were in a position of peripherality.

Although participants expressed similar motivations to engage in translation, their experience as newcomers was marked by distinct boundaries. To overcome these boundaries, however, participants faced a series of challenges that are common among newcomers who do not have a full understanding of the materials, standards and values that underpin their practice. For instance, three participants reflected on the difficulties they came across before CX was launched in 2014. Maxime (FR) highlighted that before 'the content translation tool was enabled in Wikipedia, it was really a burden to translate'. At the same time, Robin (SV) pointed out that 'the translation tool made it relatively easy to do what [they] aimed to do'. Likewise, Pau (ES) observed:

> At the beginning, I had to do it from scratch, I mean, I had to copy everything: copy and paste, let's say, from English to Spanish, and then translate it into the Spanish version. Now we have the translation tool, which, let's say, helps sometimes; not always, sometimes it does strange things. But yeah, from the very beginning, I used to translate things.

In contrast, Alex (ES), who said they had never used CX, held that their experience was conditioned by a generational reluctance to incorporate devices into their practice:

> Well, so to say, although Wikipedia is an internet thing, and a technology thing, I still behave as if I were before a typing machine, you get me? My head works the old way out, and I am just typing in Wikipedia; you get me?

Besides mentioning CX, another three participants referred to their inbound trajectory as a fun, dynamic and enriching process. Ange (FR) remarked that 'I'm not sure, I would say 'fun'? Or at least easier, or smoother, I'm not sure what I'd like to call it, but it's... And it's richer to translate an article'. In the same vein, Leslie (NL) commented that 'So, yeah, it's about expanding the encyclopaedia; for me, it's a fun thing to do'. Nik (NL) felt that part of their journey being more enjoyable was partly due to the impact of CX in configuring the practice: 'It is fun, and the current version of the tool is helpful'.

The participants' answers above reveal an overall sense of fulfilment, which is crucial to the survival and evolution of any given practice (Shove, Pantzar, and Watson 2012). Notwithstanding, it soon becomes apparent that engaging with other practitioners can put that sense of fulfilment at risk in some cases. Reflecting on their experience observing other members, Ariel (ES) argued that Wikipedia at times could be a toxic environment if someone opts to translate about seemingly controversial topics such as politics: 'I try not to take many ecological or political issues because it tends to lead to a lot of debate'. Similarly, Charlie (SV) was critical of the unwelcoming environment that, in their view, had led some inexperienced newcomers to defect from the practice:

> This [issue] has been much discussed in the Swedish Wikipedia, because there was a time when the tone was too harsh, and that discouraged newcomers. They got to know how bad of an article they had written. And, of course, you don't want to write articles if others say it's crap.

Sam's (FR) inbound trajectory was more positive. Still, their comment below corroborates the influential role that constructive feedback and mentorship from more experienced practitioners have in ensuring someone's assimilation into the shared repertoire:

> And, also, so far, I've been very lucky with the people on Wikipedia I've talked with. Because most of the time whenever I need some time or say 'okay maybe this page I translated could be included in another bigger page' or something, people have always answered like 'yes, I can help you'.

These observations are revealing of how translators in Wikipedia can, and often have, different ways of engaging with their peers while still being accountable to one another. The aid provided by materials such as automated devices will be examined, alongside aspects of community engagement, in subsequent sections of this chapter as well as in Chapter 6.

Concerning specific challenges, the participants' answers can be divided into two broad themes: a) issues related to the ST and b) technical issues. Within the first theme, recurrent problems were having to deal with specialised vocabulary, difficulties understanding the ST, deciding whether the topic covered in the ST was relevant to the target community, and finding better references when those present in the ST seemed insufficient. Regarding the second theme, participants said they

had experienced problems with some content-creation devices and unintuitive wikicode and other aspects of know-how (templates, infoboxes, categories and internal links. Table 5-3 below illustrates the major points of concern, as addressed by individual participants.

| Theme | Subtheme | Participant |
|---|---|---|
| ST-related issues | Vocabulary | Ariel (ES), Cris (ES), Ange (FR), Maxime (FR), Sam (FR), Guus (NL) |
| | Understanding the ST | Alex (ES), Cris (ES), Ange (FR), Dominique (FR), Charlie (SV) |
| | Relevance of the subject | Sam (FR) |
| | References | Charlie (SV) |
| Technical issues | Devices | Ariel (ES), Maxime (FR), Jos (NL), Alva (SV), Kim (SV) |
| | Templates | Pau (ES), Jos (NL), Robin (SV) |
| | Categories and infoboxes | Cris (ES), Leslie (NL), Nik (NL) |
| | Internal links | Jos (NL), Leslie (NL) |

Table 5-3. The participants' primary challenges in their inbound trajectory.

An area of concern for six participants, lexicon emerged as one of the top two non-technical challenges. For Ariel (ES), the primary hurdles as a peripheral Wikipedia translator were choosing a suitable title in the target language and translating technical terms used in chemistry, their field of expertise. Cris (ES), also from a scientific background, struggled with technical jargon as well, but, unlike Ariel, the difficulty in their case resided in deciding which Spanish term was adequate. According to Cris, several Spanish-speaking countries use different names to refer to the same concepts, and sometimes 'a lot of things don't have translations just

because of the weight that English has in [domains such as] the aerospace industry'. This concern addresses a potential shortcoming of the Spanish Wikipedia translation guidelines. As analysed in 4.1.5, the document advises translators to adopt a standardised version of the language. However, as Cris' comment reveals, the question becomes more complex with articles on science. This complexity was also attested by Sam (FR):

> There's also the fact that sometimes it takes time to translate some concepts which are very specific in science, so sometimes I need to go to some forums or very specific dictionaries to find the exact equivalent in French about a scientific concept.

Ange (FR) and Guus (NL) attributed the problem with vocabulary to their lack of knowledge of particular words or expressions in the ST. Maxime (FR) said that they learned to avoid translating topics they did not know much about because that minimised the risk of stumbling upon unfamiliar words. As noted in 4.2.1, this course of action is recommended in the Dutch Wikipedia guidelines, but not in the French Wikipedia documents, where priority is given to the quality of the source article over the translator's familiarity with the subject. The guidelines also come short of advice regarding how to deal with technical terminology. It is therefore up to the translator to decide the best course of action and how to meet the expectations of their readers. Ange and Guus, for example, opted to skip terms or sentences that were obscure to them when the sources they consulted were not able to give a suitable solution. Ariel (ES) adopted the same approach but as a last resort, while Sam (FR) appealed to other members of the community by posting requests on the talk page of the translated articles. Cris (ES) did not indicate how they dealt with the issues they encountered.

Understanding the ST was a matter of concern for five participants. Besides Cris and Ange, the topic was mentioned by Alex (ES), Dominique (FR), and Charlie (SV). In Alex's experience, the challenge was translating articles where the topic was difficult to grasp or the text had been written by editors whose native language was not English. Ange, referring to the subject of notable Irishwomen, commented that sometimes they had to read other related Wikipedia articles to understand the ST. Cris reported that they came across errors in the ST, which they had to correct before embarking on the translation. Charlie alluded to the 'poor' quality of the ST as their most significant difficulty:

> Well, the greatest difficulty, I should say, is when you have decided to translate an article, and then you discover that the article you're trying to translate is very bad, it's bad language, or the sources are second or third rate, you find out that some of the important things in the article are all references from a blog.

According to Charlie, another major problem was adapting the ST to a target community that may not share the same views expressed by the original editors. The Swedish Wikipedia translator referred to this issue as a 'cultural problem':

> And, sometimes, you can find a cultural problem: the person, the writer, that you're trying to translate has another way of looking at things, and you feel that 'I can't translate this as it says, because I don't really understand the writer who did it'. And, then, you also have to decide 'should I choose my own way and make the article myself, or should I try in some way to understand the article to really translate it?'

Dominique also brought up culture and ideological stance as two of their main obstacles when translating biographies into French. In this sense, their experience shares some similarities with Charlie's. Like the Swedish Wikipedia editor, Dominique noted that at times it was difficult for them to agree with the authors of the ST:

> I guess it's more difficult in these topics to fiddle in translation, because when you translate then you are bringing things into another cultural framework, so you have to defend the idea which came from the author.

Concerning content, Sam commented that they occasionally wondered whether the ST would fit well in the French Wikipedia. This doubt stems from subtle differences across Wikipedia language communities concerning notability criteria (see 4.2.1). Problems around notability may encourage some Wikipedia translators to search for additional bibliographic references to anticipate possible objections by other editors. On this ground, Charlie remarked that they checked facts on search engines such as Google whenever the ST was making unsubstantiated claims.

Apropos technical issues, four participants indicated that the lack of adequate devices posed a few challenges when they first started translating in Wikipedia. It is worth noting at this point that the devices mentioned by the interviewees were not simply restricted to machine translation. Some participants referred to the Wikipedia interface, with the emphasis on wikicode. Ariel (ES), Maxime (FR), Jos (NL) and Kim (SV) said that translation in the pre-CX era was tedious and required

editors to make a concerted effort to translate and revise the text. Along the same lines, Ariel, who joined Wikipedia in 2015, was particularly critical of senior colleagues who were still against mobilising the Visual Editor (VE), an online rich-text editor based on the WYSIWYG principle. On the difficulties of wikicode and the impact that the Visual Editor had in configuring the practice, Jos observed:

> Yeah, it's still, like, the old media before the Visual Editor. Before Vector, even. It was a very different kind of experience. The problem was that you had to do everything in a mark-up language, so basically you had to type everything from scratch, there was no tool really to help you with the translation. So the hardest thing at that time would be... The references were also not there, so that was not the problem, at the time the problem was getting the links right: so you would have to go to each article, figure out what was the corresponding article.

To conclude, the data suggest that most participants faced similar challenges in their inbound trajectory. Such challenges were related to a lack of exposure to the shared repertoire, as became apparent by the participants' complaints about unintuitive procedures, the lack of adequate devices to facilitate their task, and difficulties navigating Wikipedia's interface. Moreover, although translators of the four Wikipedia communities came across ST and technical issues, the data indicate a prevalence of technical difficulties among the Dutch and Swedish Wikipedia translators. In contrast, French Wikipedia translators, in general, did not tackle the use of devices and shifted their attention to topics related to the ST. Three of the four Spanish Wikipedia translators addressed both ST and technical problems. Considering how participants struggled with aspects of the shared repertoire in their inbound trajectory to membership, the next two sections will examine whether, and if so, how they have incorporated the documented standards that seem necessary for their initiation into the practice.

## 5.4 Wikipedia policies and guidelines

In 2.2, it was argued that practitioners subscribe to a series of standards established by their communities. As Wenger (1998) observes, such standards form part of the community's shared repertoire, alongside values and materials. Individuals can follow these standards explicitly or implicitly or even resist them (Wenger 1998). In 2.3, it was noted that compliance with the standards often involves 'knowledgeable participants able to make credible judgments' (Warde 2016, 201)

rather than strict obedience to rules. Nevertheless, there is a sense of continuity (Warde 2016) and mutual engagement (Wenger 1998) in practice.

Chapter 4 analysed the documented standards of translation practice in the form of Wikipedia policies, guidelines and essays, and examined how mutually engaged individuals negotiate them in talk pages. However, it remains to be seen whether experienced Wikipedia translators have incorporated those standards into their practice (RQ3). This section seeks to gain insight into the themes that came to the surface during the first and second blocks of the interview, which dealt primarily with Wikipedia policies, guidelines and essays. Special attention is paid to implicit and explicit compliance, as well as to rebellion (Wenger 1998). The concept will also be addressed in Chapter 6 when investigating the participants' response to the deployment of automation.

As in 5.3, the recurrent themes will take precedence in the analysis. Throughout the interview, participants were asked which, in their view, were the Wikipedia policies and guidelines that editors should follow before and during the translation process. Participants were encouraged to reflect on their practice and think about the standards they tend to prioritise when they engage in translation. As happened in the previous block of questions, answers differed considerably in some cases. Notwithstanding, the responses revealed a tendency among participants to focus on specific policies, guidelines and principles that are part of the more encompassing Wikipedia practice of editing. For instance, a majority mentioned verifiability of content (VER) as essential, followed by familiarity with the topic, the need to fill a knowledge gap, notability, post-editing, the importance of knowing local conventions, faithfulness to the ST, and the 'Five pillars'.

A reduced number of participants acknowledged that they had ignored the rules on certain occasions. Individual participants also discussed NPOV and aspects of engagement, such as not upsetting senior editors and asking for help if necessary. Table 5-4 below shows Wikipedia standards – policies, guidelines and procedures – mentioned by the participants. The documented standards of translation (policies, guidelines and essays) are not included in the table because they did not appear in the participants' first answers. As will be discussed in 5.5, a prompt was used in most cases to elicit that information. In what follows, the participants' responses

are examined, paying attention first to policies and guidelines, and then to those procedures – examples of good practice – that the participants considered relevant to their performance.

| Policy/guideline | Status | Participant(s) |
|---|---|---|
| Verifiability of content/ Acknowledgement of the ST | Policies and guidelines | All participants |
| Notability | | Ange (FR), Maxime (FR), Guus (NL), Jos (NL) |
| Knowledge of local conventions | | Cris (ES), Pau (ES), Dominique (FR), Robin (SV) |
| Five pillars | | Alex (ES), Cris (ES), Ange (FR) |
| Ignore the rules | | Cris (ES), Nik (NL) |
| Engagement | | Jos (NL) |
| Familiarity with the topic | Procedures | Maxime (FR), Sam (FR), Jos (NL), Charlie (SV), Kim (SV) |
| Filling a knowledge gap | | Alex (ES), Ariel (ES), Pau (ES), Alva (SV), Robin (SV) |
| Post-editing | | Ariel (ES), Sam (FR), Nik (NL), Kim (SV) |
| ST | | Pau (ES), Sam (FR), Kim (SV) |

Table 5-4. List of Wikipedia policies, guidelines and procedures mentioned by the participants.

### 5.4.1 Verifiability of content and acknowledgement of the ST

As discussed in 4.1.1, verifiability of content (VER) is one of Wikipedia's core tenets and overarching policies. VER encompasses a wide range of principles that go from quoting reliable sources in an encyclopaedic article to finding proper ways of acknowledging other editors' work. All 16 participants agreed on the significance of including references in the translated article. Alex (ES) stated that it

was a ‘general rule’. Pau (ES) pointed out that importing references from the ST to the TT was necessary to meet the VER criteria.

In 4.1.1, it was observed that the translation standards of the four communities stressed the importance of avoiding unsubstantiated facts. A few participants mentioned this recommendation in the interview. Maxime (FR), for instance, observed that ‘an article with no references is not supposed to be translated’, and went on to say that, as an administrator in the French Wikipedia, they had felt compelled to delete articles ‘created by people who forgot to add references’. Nik (NL), who said they were wary of translating articles with insufficient references, stated that they always included them all and tended to ‘keep them intact’, with only ‘some tweaking’ when needed.

Despite their consensus around references, participants held different opinions regarding aspects such as quantity and the obligation or optionality of consulting the sources before importing them into the TT. As analysed in 4.3.3, relying on references that one has not had access to is discouraged in the Dutch Wikipedia. A divisive issue, the matter was hotly debated in talk pages between 2009 and 2011. The outcome of the negotiation, primarily resulting from the inaction of the contesting party, was to keep the recommendation of checking the references beforehand.

Although the debate occurred almost a decade ago, the answers provided by Dutch Wikipedia participants show that some translators still resist this piece of advice. Leslie said that they did ‘not have a strong opinion on it’ but was inclined to agree for the sake of transparency. Guus, Jos and Nik declared that they had ignored the recommendation. Nik said that they did not consult all the references of the ST because they trusted the original authors. Jos was outspokenly critical of their colleagues, noticing that ‘a very small group of people [in the Dutch Wikipedia] care a lot about a very small set of rules’. Similarly, Guus addressed the subject directly and asserted that reviewing all the sources was onerous:

> Another discussion which is always there is the sources, because if you translate an article... At least when I translate an article I usually don’t go into the sources that are used for that article, I just translate it, and I try to assess if it’s true, if it seems correct, but I don’t check it. This is a discussion which always occurs because some people are very strict and say you should check all the sources yourself... Yeah, in that case, I would if that were a real policy; this is only a

> discussion, it's not an actual policy, but it's always in discussion, and I think if I have to check the source myself, I would rather write it in my own words, if I have to do so much work on the sources then I won't translate.

As far as the number of references is concerned, there were slight differences in criteria among participants. Alex (ES) opined that the quality of an article is assessed by the number of references. Jos and Leslie, however, belittled the weight of the sources. Reflecting on their experience translating biographies from English into Dutch, Jos commented that the number of references in the ST was not a criterion for them. They added that their 'assumption is that if there are no references [in the ST], it is probably because the article is too old, and if it's been in that shape for so long it's probably a fine article'. Leslie noted that English Wikipedia editors tend to 'have every sentence sourced'. In Leslie's view, this is not common practice in the Dutch Wikipedia; it is not part of the community's shared repertoire. For this reason, they thought it was correct to 'delete some references' when they translated.

When it comes to additional references, the participants' answers show some divergence as well. Pau (ES), Ange (FR), Sam (FR) and Jos (NL) said that they were reluctant to search for references in their language. Ange mentioned that there are not many sources available in French on the subject of notable Irishwomen, their topic of interest at the time of the interview. Likewise, Jos observed that there is a scarcity of publications in Dutch. Pau posited that if an article is not available in the Spanish Wikipedia, it is because 'maybe that topic is not very popular or very common' in the Hispanosphere. Sam indicated that they rarely added references in French to the TT because most scientific publications are in English. In contrast, Leslie (NL), Nik (NL) and Charlie (SV) said that they always endeavoured to find additional references in their language.

Closely related to the verifiability principle is the acknowledgement of the ST, as studied in 4.1.2. Since most Wikipedia translation guidelines strongly recommend crediting the ST, one part of the interview tackled this issue directly by asking participants if, in their opinion, acknowledging the source was essential. As a follow-up question, participants who responded affirmatively were asked to elaborate on the procedures they followed.

The overwhelming majority of participants reported having credited the source of the translation. Nevertheless, the rationale for this course of action varied across interviewees. For most of them, crediting the ST was neither mandatory nor strictly enforced. Instead, crediting the source was regarded as good practice. Pau (ES), for example, stated that reusing content without indicating its origin was unethical. Dominique (FR), for whom crediting the ST is a 'legal requirement', remarked that doing so was not always practical or accurate because 'one version in one language might or might not reflect the same content that there is in the native or whatever language you started with'. Similarly, their colleague Maxime (FR), compliant but critical, disapproved of this procedure on the grounds that Wikipedia articles change over time:

> I don't know, but I did that for a bit when I helped a guy who translated an article from English Wikipedia into French Wikipedia and then the articles have diverged and are totally different nowadays. And we still have this credit that says 'this article is a translation from the English article'... It's not the case at all anymore; you may find some paragraphs that are the same on both sides, but we have entirely rebuilt the article. Everything changed, the structure changed, we have written some elements, so... I don't see the point of attributing an article as a source, as it's not the source anymore, it's not... Well, let's say you have 15% of the article that is from the English Wikipedia now... What's the point?

While most participants concurred that crediting the ST was required, their answers revealed some disagreement about the means to achieve that end. This dissent was already observed in 4.3.2, where the data showed that the ambiguity around translation templates and their usage had triggered a lengthy debate. The interview data corroborated the lack of a standard procedure, partly attested by an overlap in the participants' responses. When asked about how they usually credited the ST, six participants stated that they did so automatically via CX. The device generates the link to the ST in the revision history of the translated article. Another six participants said they usually added the template at the bottom of the translated article within the 'references' section. Four interviewees argued that it was better to credit the ST on the talk page of the translated Wikipedia article because of its visibility to readers. Finally, two interviewees, Cris (ES) and Kim (SV), were of the view that crediting the ST was unimportant.

### 5.4.2 Notability

As Table 5-4 illustrates, half of the participants identified notability and post-editing as two essential criteria they apply when translating an article. Notability is a precondition for an article to be accepted in Wikipedia. Nonetheless, sometimes what counts as notable or relevant hinges on the editors' interpretation of a given language community. In 4.2.1, notability appeared as one of the criteria for article selection. As such, it features in the French Wikipedia guidelines ('Aide:Traduction' 2020) alongside verifiability (VER) and neutrality (NPOV).

Four participants addressed notability as a criterion. Two French Wikipedia participants, Ange and Maxime, came up with the concept during the interview. According to Ange, notability is their first criterion. The editor commented that, on occasion, they felt that it was necessary to abort the translation after realising that the article they had chosen was likely to be considered irrelevant in the French Wikipedia. When prompted to elaborate on that statement, Ange attributed that capacity for distinguishing between notable and unrelated to an implicit knowledge acquired through performance:

> Because if it's not notable enough the article could be deleted and you've worked basically for nothing, so clearly it's the first thing I look, and it's quite easy... I've been here for a long time, so now I know, and I don't need to look too much at it. I look at the beginning, I make sure the article is reasonably admissible and notable, and then I can move on. That's the first step, and the most important one, I would say.

In a similar vein, Maxime also referred to notability as a standard that varies from one Wikipedia community to another. Notwithstanding, the participant said that 'inclusion criteria' – as they called it – can be met in the French Wikipedia if the translated article is well-sourced (VER). On speaking about notability in the Dutch Wikipedia, another interviewee, Guus, observed that the standard is 'very subjective' and disputed. Like Maxime, Jos tied notability to 'a wealth of references' in the article. Therefore, it seems apparent that for a few participants, notability and verifiability are intertwined in Wikipedia.

### 5.4.3 Knowledge of local conventions

In the previous subsection, it was noted that standards such as notability vary across Wikipedia communities. As argued in chapters 2 and 4, these differences can be explained by the existence of distinct shared repertoires across Wikipedia language communities. According to four interviewees, this lack of awareness about local conventions has the potential to create a barrier or boundary between peripheral translators and senior translators. Cris (ES), for instance, reported that they knew editors whose translated articles had been harshly criticised by their peers for not following Spanish Wikipedia naming conventions.

Pau (ES) stated that '[Wikipedia] communities think differently, and sometimes they mean things in a different way'. For Dominique (FR), translators should attempt to adapt the ST to the target Wikipedia community, being mindful of the cultural context. Robin (SV) declared that a good Wikipedia translator ought to be familiar with their community's manual of style, paying attention to 'slight differences in how [editors] structure articles in different language versions'. To illustrate this point, Robin gave an example of birthday templates and how they are used in the Swedish Wikipedia:

> For instance, in the English language Wikipedia they always place birthday templates first and the date of death within brackets, while in Swedish, it's commonly separated by a comma instead. So, those are kind of small things that I think if a new user wants to be appreciated for the work that they're doing, they should perhaps also read... Or just look into how other articles are looking to see the structure and the more aesthetic parts.

### 5.4.4 The 'Five pillars' and other policies and guidelines

Three participants referred to the 'Five pillars', a set of core Wikipedia policies ('Wikipedia:Five Pillars' 2021). These tenets were discussed in 2.4.1 and 4.1. As noted in the previous chapter, the policy states that a) Wikipedia is an encyclopaedia; b) Wikipedia is written from a neutral point of view; c) Wikipedia is free content that anyone can use, edit, and distribute; d) editors in Wikipedia should treat each other with respect and civility; e) Wikipedia has no firm rules. These fundamental principles encompass a wide range of policies and guidelines aimed at ensuring a

healthy and sustainable work environment. Therefore, as mutually engaged individuals, Wikipedia editors are expected to comply with these rules.

Alex (ES) commented that they '[gave] importance to the Five pillars'. Cris (ES) asserted that compliance with the policy was obligatory: 'For me what you always have to follow no matter what are the 'Five pillars'. Likewise, Ange (FR) observed:

> What we call them in French, we have the... How do you call them, the five principles: *'principles fondateurs*' in French... Five pillars in English; those are not negotiable. Rules are very important.

The fact that only three interviewees mentioned this set of principles is illustrative of the explicit/implicit duality of practice. As engaged members of their communities, translators – and editors in general – are expected to adhere to these basic tenets, even if they are not able to list them. For example, creating non-encyclopaedic content, violating neutrality and being systematically rude to others could prove detrimental to someone's inbound trajectory to membership in Wikipedia. By the same token, mention of particular policies and guidelines, when explicitly prompted, may mean that these are the most salient ones for those individuals at that specific moment.

Other policies and guidelines also emerged during the interviews. As illustrated in Table 5-4, a minority of participants raised further issues. Cris (ES) and Nik (NL) held that at times it is correct to ignore the rules. However, according to the interviewees, disregarding the standards does not mean that translators have leeway to do as they please. On the contrary, it is an appeal to common sense or implicit compliance. Wikipedia's excessive bureaucracy has been criticised by some scholars (O'Sullivan 2009), whereas the issue of documentism in communities of practice has been addressed by Wenger, McDermott, and Snyder (2002). Wikipedia itself acknowledges this flaw in its fifth pillar, providing some reassurance to editors with the message that the encyclopaedia does not have firm rules.

Another participant, Jos (NL), tackled aspects of community engagement. The translator, who had previously been critical of a reduced group of rule-bound colleagues, went a step further and denounced what they referred to as 'dinosaur privilege' in Wikipedia. As a senior editor registered in 2005, Jos outspokenly criticised the divide between full-fledged editors with a sense of entitlement and disempowered peripheral users:

> In Wikipedia, you have some kind of equivalent to white privilege, where you have basically like 'dinosaur privilege', let's call it that. If you are a dinosaur in Wikipedia, you can act whatever way you want, and you can get away with it. That is not true for a new contributor. So if you if you're a new translator, like, if you have a very new account, I would definitely advise you not to do that, and I would advise you to make sure you don't upset people, rather than follow the policies because they are good policies. Because upsetting people will set you back so much time and so much effort that it's just not worth it.

Finally, Dominique (FR) touched on editorial neutrality. Like verifiability and notability, neutrality (NPOV) is an overarching policy of Wikipedia. This principle was discussed in Chapter 2, particularly in 2.1.1, and it was mentioned again in Chapter 4 as one of the criteria for selecting a suitable article for translation (see 4.1 and 4.2.1). As an integral part of the 'Five pillars', NPOV features prominently in some translation guidelines such as *Ayuda: Cómo traducir un artículo* [Help: How to translate an article], *Aide: Traduction* [Help: Translation], and *Wikipedia: Översättningsrekommendationer* [Wikipedia: Translation tips]. During the interview, Dominique stated that neutrality was a policy that they pursued in their practice, not without struggle:

> So, given that, I would say neutrality is an interesting point; because it gives an idea that, while it's already something that can be debated: 'what does that mean? To be neutral', on the other hand, it can provide a good guideline even if you want to debate what is neutral; you can do that for each community, for each language community at least, because of course there's also the problem that English or French, for example, are languages that are used in many countries, which will have their own cultures and their own points of view, so it can also lead to conflicts, so conflict is not... It's not something that you don't want to avoid, but you have to be sure that you are doing it to create or use some process that will enable to overcome these kinds of issues.

### 5.4.5 Familiarity with the topic

As analysed in 4.2.1, being versed in the subject of the TT is highly encouraged in the Dutch Wikipedia translation guidelines. The page strongly recommends avoiding articles that are outside one's field of expertise. In the previous stage of the analysis, it became apparent that this piece of advice forms part of the Dutch Wikipedia community's shared repertoire. In contrast, it was noticeably absent – at least in writing – from the other Wikipedia communities under investigation.

The interview data suggest that, while the standard is not mentioned explicitly in other communities, it is present tacitly in some participants' practices. For example, Maxime (FR) said that they tend to 'pick topics [they are] familiar with'. Sam (FR) explained that for them, it 'would not feel okay to translate an article from another domain of knowledge, even if the English [Wikipedia] page is well-sourced'. In a similar vein, Charlie (SV) indicated that their most important criterion for selecting an article was having at least a basic understanding of the subject. Kim (SV) held that 'some sort of preconception of what the concept is about is as important as knowing something about the subject'. Jos (NL) also highlighted that familiarity with the topic of the TT was of 'paramount importance' to them.

Jos was the only Dutch Wikipedia participant who brought up the issue of familiarity. Nevertheless, attributing their preference for article selection to a knowledge of the translation guidelines would be an erroneous assumption to make at this stage. As will become apparent in 5.5, most participants were unaware of translation standards within their communities, and those who did know about their existence declared that those pages were irrelevant to their practice.

A more plausible explanation to account for the participants' predisposition to opt for familiar topics is found in their motivations to engage in translation and their position as brokers. With personal or professional interest being the most prevalent reason to become translators in Wikipedia (see 5.2), it would be reasonable to conclude that some familiarity with the subject is implied. At some point, this criterion was negotiated by a group of editors and became part of the Dutch Wikipedia shared repertoire as an explicit value of the practice, while it remains a matter of personal choice (implicit, non-documented) in the other communities under investigation.

As regards brokering, translators in Wikipedia are not bound to their language community. They have access to a wide range of perspectives and may even participate in other language communities or sister Wikimedia projects such as Commons or Wikidata. Therefore, a few participants in the study may have been exposed to different communal standards, which they subsequently incorporated into their practice. As noted by Wenger (1998, 109), brokering is not unusual in communities of practice, where some members introduce standards that are not

necessarily known to others. Such standards are likely to spring from multi-membership: that is, belonging to more than one community of practice. Chapter 6 will expand on this idea when tackling the participants' views on the mobilisation of devices and ascertaining the impact of Wikidata on translation.

### 5.4.6 Filling a knowledge gap

The previous findings are reinforced when one examines the second recurrent criterion for article selection, as stated by the participants. Filling a knowledge gap was among one of their motivations to engage (5.2), and yet, for some, it played a role when it came to choosing an article to translate. The so-called 'red links' (interlinks) are good indicators of Wikipedia articles that are missing. Three of the four Spanish Wikipedia participants referred to this fact during the interview. For Pau, their selection of articles was sometimes arbitrary. According to the interviewee, when they translated their first entries over a decade ago, they initially looked for missing articles that were available in other encyclopaedias such as Encarta and Britannica. Their colleagues Alex and Ariel also targeted missing articles. Ariel's answer elucidates the role of red links in determining which Wikipedia articles could be translated:

> When you start translating something, pages that do not exist in the language start popping up, so sometimes it turns into a ladder: you start on a topic and end up translating articles about an entirely different subject. You noticed they were small articles, and in order not to leave them full of empty links; sometimes I do that: I start and see where it takes me.

Two of the Swedish Wikipedia participants had criteria akin to those described above. Robin said that they had encountered red links when translating other articles. Once again, the urge to fill a knowledge gap is what prompted this participant to translate related Wikipedia articles. The interviewee observed that they 'would perhaps do a translation of an article…to make sure there's not a red link in the other article [they're] building'. Likewise, fellow translator Alva stated that they often searched for missing articles 'that are needed [in the Swedish Wikipedia]'.

The translators' responses reveal that their criteria are aligned with Wikipedia's values. As explained in 4.1, the user-generated encyclopaedia welcomes and encourages contributions from volunteers. The platform's slogan and welcoming

message, '[the] free encyclopaedia that anyone can edit', promotes this kind of philosophy. At the same time, the data suggest that the participants' motivations often coincide with the benchmark they have for translation. To put it more concretely, it is not always possible to draw the line between what prompts editors to translate and the criteria they apply for article selection.

The fact that participants from different communities show the same alignment with Wikipedia's principles lends support to the postulate that the encyclopaedia is a constellation of communities of practice. This statement is further attested by the priority that most participants give to overarching policies and procedures such as verifiability, notability, and post-editing.

### 5.4.7 Post-editing

As regards post-editing, a quarter of those interviewed suggested that it was essential for translation. The four participants were unanimous in the view that post-editing was critical when mobilising automated devices. Ariel (ES) said that they always proofread and edit the TT thoroughly after importing it from Google Translate. Sam (FR) recognised that post-editing an MT-generated text could take time, but they argued that it is a crucial step 'to keep [an article] as good and accurate and relevant as possible'. Nik (NL) went a step further and asserted that publishing an unrevised MT-generated text is forbidden and that without proper post-editing and amendments, the article is likely to be deleted by a Wikipedia administrator. Kim (SV) explained that as a translator in Wikipedia, they felt that post-editing should be done out of respect for the readers:

> Wikipedia is a part of the internet and, as such, we make all information about the world available as we possibly have the time to do. Still, at the same time, we do not have to give them false expectations and write things that we don't know if they're really true. So, it's some sort of idea when you translate an article or write an article in general, that this will not be over within a minute or even an hour, because it can take days. It can take ten or fifty or hundreds of edits before the article looks like what you expected from the beginning because you didn't know how hard it was to get it to this.

### 5.4.8 Faithfulness to the ST

As with post-editing, faithfulness to the ST is neither a policy nor a standard in Wikipedia. Unlike post-editing, however, being faithful to the source is not always expected in the online encyclopaedia. If anything, translated articles may evolve over time to a point where the source text is no longer easily recognisable (Jones 2017; Shuttleworth 2018). In the sections above, the data revealed that some participants took liberties with the text, omitting content that, in their view, was not relevant to their community. Others opted to skip obscure terms or expressions altogether. Despite these findings, three participants (see Table 5-4) said that faithfulness to the ST was relevant to their practice. Pau (ES) mentioned it as their priority: 'Well, the first I would say: you need a faithful translation into Spanish'. Sam (FR) claimed that the TT should aim to resemble the ST in content:

> So, first, I would say, 'don't damage the original page you're translating'. If you wanna paraphrase, that's okay as long as you say the same sense that the original sentence you're translating wanted to mean.

Kim (SV), who was not against faithfulness to the ST, observed that in some instances, translators should be able to adapt particular terms or expressions to suit the target community's cultural context:

> Well, yes...Well, that's like translation in general, because you have to be true to the target language; you cannot translate a thing that's called one thing in English and Catalan and expect it to be called exactly the same thing in Swedish, or as a direct translation: you have to know what you're talking about, and that's a general thing.

Thus far, the data have shown that most participants have incorporated, albeit implicitly, some of the most significant Wikipedia's core policies and guidelines into their practice. During the interviews, it became apparent that participants were in tune with the primary policies of the encyclopaedia. Some contested a few aspects, such as crediting the ST and importing references. Others were critical of the emphasis that a small number of editors place on standards. Such criticism could be interpreted through the lens of Wenger's (1998) concept of rebellion, which, as explained in 2.2, is not necessarily an antagonistic position. Far from it, challenging specific conventions and drawing attention to particular behaviours is usually done from a place of high commitment (Wenger 1998). In this case, questioning

established standards of practice in the shared repertoire may pave the way for meaningful changes over time. On the other hand, passive conformity may lead to the perpetuation of specific values, as observed, for example, in 4.3.3. The following section will return to the subject of rebellion in practice when analysing the participants' knowledge and assimilation of the translation standards set by their Wikipedia communities.

## 5.5 Knowledge and incorporation of translation standards

When explicitly asked about translation standards, most participants were unaware of their existence or downplayed their significance. Eleven participants regarded Wikipedia translation standards as unimportant. Of these, some were under the impression that most translators did not consult them. Others went on to say that the documents are not easy to find. Among those who minimised the relevance of standards, three thought that the standards were outdated and questionable. In contrast, only two participants, both from the Swedish Wikipedia, indicated that they were familiar with the essay. One of them was even able to comment on aspects of the page. None of the Spanish Wikipedia participants explicitly mentioned the guidelines. In the remainder of this section, the participants' answers are examined in more detail.

Three of the seven participants that gave little weight to the standards were from the French Wikipedia. For Ange (FR), translation guidelines 'could be useful, but more as help documentation that's always very specific and doesn't always apply to what you're doing'. Their colleague Maxime was more critical. According to the participant, 'help pages'– as they called it – were obsolete in the French Wikipedia. Nevertheless, Maxime noted that this obsolescence could not merely be narrowed down to outdatedness. Instead, the problem resides in that editors, in general, ignore the rules or learn them elsewhere in Wikipedia:

> Something communities haven't thought about is the fact that people, you know, don't read the documentation. That's the same thing when you have a new TV, for instance, a new appliance at your home: you put it on the table, you plug it [in], and then you try to have it working. That's the same thing when people use Wikipedia: they don't read the manual.

Maxime also noted that, in their community, experienced translators often contribute to the guidelines in good faith to help newcomers or peripheral practitioners. Nonetheless, in the participant's view, such endeavour is of little avail. Their colleague Sam, who overtly recognised that they were among those who had not 'read and integrated [the guidelines]', believed, however, that they had sufficient knowledge of Wikipedia rules. Unlike Ange and Maxime, Sam regretted not having assimilated the translation guidelines of their community: 'I'm doing some good work, but you're right, I should still take time one day to read them carefully'.

All four Dutch Wikipedia translators declared that they had not paid much attention to the guidelines. Guus pointed out that the Dutch Wikipedia page does not constitute 'real or formal guidelines'. Despite this fact, Guus observed that 'some people are always hammering on [the topic]'. Fellow translators Leslie, Nik and Jos mentioned that they followed common sense rather than specific standards. This disregard for the guidelines suggests that, at least in their case, long-time engagement in the same practice is underpinned by implicit knowledge of what works well in their community. The implicitness of practice is ascertained in Jos' comment below, which also encapsulates the capacity that experienced translators have to make reasonable judgments about their practice:

> I've been doing this for so long I have no idea what the policies exactly are these days. Like, I just follow common sense mostly. Dutch Wikipedia is not a very policy-heavy Wikipedia in the sense that we have probably a lot of policies, but people don't... Some people care a lot about them, but most people don't care that much.

Moreover, the interviewee brought to the surface an inherent problem of Wikipedia: documentism. The community disorder, which was analysed in Chapter 4, tallies with Maxime's observation above regarding the inefficacy of help pages in their community. Jos referred to the recommendation on references examined in 4.3.3, which prompted a prolonged debate in the talk pages. The outcome of the negotiation (or the lack of it, as no discernible agreement was reached) favoured the consultation of sources before being imported into the TT. Although the advice has not changed in almost a decade, Jos argued that the seeming lack of objection from other members of the community does not necessarily result in compliance or conformity but, rather, in tacit rebellion:

> So, for example, there is a policy that says that if you include references you have to open every reference and check it and actually verify that the information is there. Nobody does that, let's face it, there is just no way that people actually go to the library, look up the book, and then check the book. There is just no way.

Like Maxime, Jos indicated that there is a gap between written standards – what editors are advised to do – and what is achieved in practice. Regardless of this deviation from the standards, translators seem to succeed in their practice. While further research is necessary to assess the impact of guidelines on the community as a whole (see 7.2), the participants' responses reveal that those documents do not inform their practice.

This dissociation was also apparent in the answers provided by the four Swedish Wikipedia translators. In this instance, however, the lower status of essays – as opposed to policies and guidelines – may explain why none of them attached much importance to *Översättningsrekommendationer* [Translation tips]. Robin said that, as far as they knew, Swedish Wikipedia lacked specific policies and guidelines tackling translation. On speaking about the 'tips', their colleague Charlie commented that they were not aware of any updates to the page because they did not read it frequently. Similar to what other participants expressed, Kim remarked that the document in their community is 'not the most important rule [they] have, although it can be useful'. Finally, like Charlie, Alva indicated that they had consulted the recommendations. However, the participant claimed that the 'tips' were outdated, and they attributed that stagnation to the toilsome negotiation process investigated in 4.3:

> Sometimes we have tried to change the recommendations, and there have been extremely long discussions. Well, I'm going to restart this debate on how we're going to do it because I think we should add data. Exactly how to do it is a whole another issue.

To conclude, the analysis of the data indicates that there is widespread agreement among participants concerning the little relevance that translation standards have on their activities. Despite not being familiar with these standards, most interviewees appear to be in tune with the core principles that underpin editing and translation in Wikipedia. The relatively lower status of guidelines and essays compared to policies may explain why adherence to the advice given in those documents is left to the discretion of the translators. Therefore, as with any

recommendation or help page, individuals may choose to integrate or dismiss particular aspects of the document. This liberty contrasts sharply with the more enforceable nature of editing policies, as analysed in 5.4. The fact that participants were able to mention a few of them during the interview reveals that they play a more significant role in their practice. This conscious reliance on editing policies over translation guidelines also lends support to previously formulated claims that editing and translation in the user-driven encyclopaedia form a continuum (Shuttleworth 2017; Jones 2018b).

## 5.6 Chapter conclusion

This chapter set out to investigate if experienced Wikipedia translators of the four communities had incorporated the standards into their practice. The interview dataset revealed that most participants that took part in the study had either a personal or professional interest to engage in translation. Their responses also suggest that some of them faced common challenges in their inbound trajectory from peripheral to full-fledged practitioners. Among the hurdles that the participants had to overcome along their journey were ST-related and technical issues. For some participants, understanding the ST vocabulary was essential. For others, unintuitive wiki code, devices and templates posed problems.

Regarding the incorporation of documented standards of practice, the analysis found that, in general, participants had a robust knowledge of Wikipedia's core policies and guidelines. Several interviewees emphasised verifiability of content, notability, and being familiar with local policies as crucial values in the translation process. Another important finding is that in some cases, it is not possible to disentangle motivation from article selection criteria. This fact became apparent when some interviewees indicated that they tended to choose articles within their field of interest or expertise.

Despite being familiar with essential Wikipedia policies and procedures, the majority of participants were not acquainted with the translation standards negotiated in and approved by their communities. The data show that participants shared a widespread scepticism as to the usefulness of these documents. A small number of interviewees was also critical of the importance that individual editors

attach to them. Regardless of these views, the findings reveal that the practice of the Wikipedia translators interviewed in the study seems to be informed by overarching standards that are commonly found in other forms of editing. Although the translation standards provide documentation that could help peripheral newcomers in their inbound trajectory towards full membership, these experienced Wikipedia translators appear to have incorporated some of the principles outlined in the pages with little, if any, exposure to them.

The next chapter focuses on the participants' mobilisation of automated content-creation devices such as bots and CX. The second part of that chapter will also investigate how devices have contributed to the evolution of translation practices in Wikipedia, with the focus on the past five years.

# Chapter 6. Devices as configuring elements of practice

This chapter sets out to examine the materiality and evolution of translation practices in the four Wikipedia language communities under investigation. In particular, it aims to answer the last of the research questions, RQ4: 'How and to what extent have automation and metadata contributed to changes in translation practices in Wikipedia over the last five years?' To this end, as in Chapter 5, the analysis will draw primarily upon data elicited from the semi-structured interviews undertaken with 16 experienced Wikipedia translators.

The previous two chapters studied how documented standards of practice are regulated and negotiated in the four Wikipedia communities (RQs 1 and 2) and the extent to which 16 translators have incorporated them into their practice (RQ3). Although the Spanish Wikipedia debate on bot-generated content was examined in 4.3.1, the analysis only tackled the views of a reduced group of highly engaged editors in 2008. Likewise, in the remainder of Chapter 4 and Chapter 5, the analyses offered insight into some editors and translators' attitudes towards MT and bespoke devices such as Wikipedia's CX. However, questions regarding the 16 participants' usage of, and views on, automation were not addressed. Against this backdrop, this chapter seeks to gain a better understanding of how multilingual content-creation devices such as bots and CX have contributed to configuring the practices of Wikipedia translators. In doing so, special attention is paid to the evolution of translation in Wikipedia resulting from automation usage.

As discussed in 2.4, materials can be conceptualised either as constituent or configuring elements of practice (Nicolini 2012; Shove, Pantzar, and Watson 2012; Shove 2017; Olohan 2021) or as external mediators in their performance (Schatzki 2003; Schatzki 2010). In the same section, it was argued that this thesis would follow Shove, Pantzar, and Watson's (2012) view of materials as integral elements of practice on the grounds that these are required for the enactment or performance of practices in online sites such as Wikipedia. Specifically, the analysis to be conducted in this chapter draws on Shove's (2017) concept of 'devices', which she

considers to be one of the three integral components of practices, the other two being resources and infrastructure.

As stated in 2.4, infrastructure, devices, and resources exist in a relation of continued interdependence (Shove 2017, 160). For instance, to perform the practice of translation in Wikipedia, it is necessary to have an electricity-powered (resource) computer (device) with Internet access (infrastructure). Although translation in Wikipedia and elsewhere (Olohan 2021) depends on the deployment of resources and infrastructural arrangements as much as on devices, the latter 'are implicated in defining [and potentially shaping] the practice itself' (Shove 2017, 159). As Vrandečić (2020) observes, regular software updates in Wikipedia have undoubtedly influenced how volunteers edit and engage with one another. For instance, the practice of editing in the online encyclopaedia underwent significant changes when the launch of the Visual Editor in 2013 resulted in a notable shift from traditional wikisyntax to a more user-friendly WYSIWYG format where editors no longer had to learn intricate coding.

Despite the impact of bespoke technologies on Wikipedia editing, research on the use and role of cross-lingual creation tools in translation practices is still lacking. Except for McDonough Dolmaya's (2017) study on (the lack of) translation policies in Wikipedia, where devices such as CX are briefly mentioned, the weight of automation in relation to the generation of multilingual content in the online encyclopaedia has generally been overlooked. Consequently, the aim of this chapter is threefold. Section 6.1 intends to shed light on the impact of external or non-Wikipedia devices on the participants' practices. Then, 6.2 moves on to deepen the understanding of how Wikipedia translators mobilise and view automated devices, including, but not limited to, bots and CX. Upon examining automation usage, 6.3 seeks to ascertain the extent to which the incorporation of such devices has transformed the translators' practices, with the emphasis on the last five years. In doing so, attention will be paid to potential differences between the four Wikipedia communities.

## 6.1 External devices in Wikipedia translation

This section starts by examining the 16 participants' use of devices such as dictionaries, MT and databases in their translation. As stated in 1.1.1, previous research on Wikipedia translation found that editors that engage in this activity often resort to dictionaries, MT, spell checkers and glossaries to enhance the quality of their work (Laxström, Giner, and Thottingal 2015). Notwithstanding these efforts by editors, research by McDonough Dolmaya (2015) has shown that transfer and grammatical errors in translated Wikipedia entries are relatively frequent and persist over time, mostly stemming from MT usage.

As became apparent in 5.3, devices such as CX, bots and the Visual Editor form part of the encyclopaedia's shared repertoire, even if some editors opt not to use them. Other materials such as talk pages are more difficult to elude. As this section will show, the devices that aid Wikipedia editing may also come from elsewhere on the Internet. Thus, to gain insight into which materials feature more prominently in the practice of Wikipedia translators, the 16 participants introduced in 5.1 were asked the following question during the interview: 'What resources do you normally use when you translate?' (See Appendix III). It is worth noting that the term 'resources' in this context is used in its lay sense, as a synonym of tools.

Two participants, Cris (ES) and Robin (SV), reported not using any devices external to Wikipedia. When asked to elaborate on their answer, Robin said they were unable to think of any option. Cris stated that '[they] simply guide[d] [themself] using [their] own language knowledge'. A third participant, Ange (FR), who recognised using online dictionaries on occasion, added that after performing a few translations on a similar topic in Wikipedia, they felt they had to resort to internal devices such as CX.

On the surface, the answers above suggest that a minority of participants perform their practice without the aid of external devices. However, such responses should be considered with caution. For instance, the fact that Robin failed to give an answer during the interview does not necessarily mean that they do not resort to devices. Equally, in 5.3, Cris (ES) admitted having problems with technical jargon and understanding the ST, especially when they were a beginner. Therefore, it could be

possible that some Wikipedia translators' reports of device usage are necessarily partial and, at times, contradictory. This contradiction is further attested by Cris' ambiguous comment below:

> Even though I have several English-Spanish dictionaries, I've never used them to translate articles… and, well, as I said earlier, when I have doubts about what the correct way to translate a word or phrase is, I usually turn to tools.

For the rest of the participants, the mobilisation of at least one device was common. Table 6-1 below outlines the most relevant devices described by the interviewees.

| **Translation-aid devices** | **Participants** |
|---|---|
| Online dictionaries | Alex (ES), Pau (ES), Ange (FR), Maxime (FR), Sam (FR), Leslie (NL), Nik (NL), Alva (SV) |
| Translation database | Alex (ES), Ange (FR), Maxime (FR), Sam (FR), Leslie (NL) |
| Machine Translation | Pau (ES), Sam (FR), Guus (NL), Jos (NL), Nik (NL) |
| Search engines | Ange (FR), Leslie (NL), Kim (SV) |
| Printed dictionaries | Ange (FR), Guus (NL) |

Table 6-1. List of the most frequent internal devices mobilised by the participants.

As Table 6-1 above illustrates, 50% of the participants in the study observed that online dictionaries were regular devices in their practices. Usage of printed dictionaries was less substantial. Overall, only two participants acknowledged using offline materials, and the vast majority indicated a clear preference for online devices. These findings show some similitude with those disseminated by previous studies (Laxström, Giner, and Thottingal 2015), especially regarding the tendency, among some Wikipedia translators, to rely on online sources, including machine translation (MT). In the following paragraphs, the participants' responses are analysed in greater length to show that the mobilisation of internal devices plays a significant role in configuring translation practices in the encyclopaedia.

### 6.1.1 Dictionaries and databases

The use of dictionaries, whether online or offline, was more prominent among French and Dutch Wikipedia translators, with three out of four participants in each community indicating that they had mobilised these devices. In the Spanish Wikipedia, two participants mentioned using dictionaries, and in the Swedish Wikipedia, only one interviewee declared consulting them. These differences, while minor, are quite revealing of the significant role that devices like dictionaries seem to play for eight participants in the study. However, as the answers below will demonstrate, there were differences among participants concerning the frequency of usage.

The three French Wikipedia translators – Ange, Maxime and Sam – cited similar reasons for using dictionaries. For Ange, online dictionaries such as WordReference were their frequent choice because they provided them with context. Meanwhile, for Maxime and Sam, the shortage of examples in specific dictionaries meant they often had to resort to other devices such as the Linguee, Glosbe and Termium databases. Likewise, the Dutch Wikipedia participants Leslie, Nik and Guus stated that other devices usually accompanied their usage of dictionaries. In Leslie's case, the translation database Reverso Context provided the best alternative for them. In contrast, for Nik, Google Translate allowed them to retrieve a wide range of synonyms for the same lexical item. Guus, the only participant with Ange who indicated using printed dictionaries, commented that they also obtained information from books.

Along the same lines, the Spanish Wikipedia interviewees Alex and Pau combined dictionaries with books and databases. According to Alex, they approached sites like Linguee carefully because 'it's a comparison of two corpora and one must go around searching the right usage across languages'. Pau, on the other hand, turned to Dictionary.com and LEO whenever they could not find a term in their home library. Swedish Wikipedia editor Alva said that, for them, dictionaries were more useful not during the translation process but for revising translated content.

As shown in Table 6-1, five of the eight participants who reported using online dictionaries also utilised databases. Besides the already mentioned Linguee, Glosbe

and Termium, translators like Alex (ES) identified as frequent users of the EU's terminology database IATE. Sam (FR), a user of Termium, shared that they became aware of this device after 'talking with people who are in translation'. In a similar vein, Leslie (NL) put their trust in the advice received from other colleagues and began using Reverso Context because of its efficacy in 'finding a proper nuance for difficult words'.

### 6.1.2 Machine Translation

Half of the participants who resorted to dictionaries and databases also emerged as users of machine translation (MT). Of note is the fact that three of the four Dutch Wikipedia interviewees are within this group. As stated in previous chapters, Dutch Wikipedia is known for its past reliance on automation, with over 50% of its articles having been bot-generated. While the data suggest that MT may have gained acceptance among Dutch Wikipedians, it is still early to draw a conclusion, even more so when none of the Swedish Wikipedia participants are in this category.

The three Dutch Wikipedia interviewees who feature in Table 6-1 mentioned Google Translate as their habitual device. Guus, who recognised that the Google Translate is not yet 'optimal', noticed, however, that it has been improved in the last years. For Guus, such advancements have enabled them to use the device 'to translate whole sentences'. This view on the progress of MT is shared by Sam (FR), who went from avoiding automation to fully integrating it into their practice:

> I would say most of the time now. At the beginning, I wasn't [a user of automation], and it was taking two to five times more time for me to translate the page. Now, I think, also the automatic translator has improved. I'm usually using one called DeepL.

Another example of the impact of external devices in configuring translation practices in Wikipedia is found in Jos' (NL) response below. The Dutch Wikipedia editor, who cited Google Translate as their only frequent external device, observed that in their case, MT was necessary to bridge the language barrier:

> If I'm translating from a language I'm not super familiar with, I will actually take the original article, Google Translate the whole article, and then re-write the article from scratch using both versions and also all references that are linked so that I understand the context properly.

Their colleague Nik, whose reasons for using MT were outlined above, acknowledged that despite the affordances of automation, devices such as Google Translate still 'make mistakes because they don't get the context right, so one has to work on that to find the more correct translation or more natural translation'. This last remark is in tune with previous answers provided by the participants on the critical use of automation, particularly in 5.3. Section 6.2 will expand on this topic when examining the 16 participants' views on, and experience with, bots and CX.

To conclude this section, it is apparent that for most participants, the mobilisation of external devices helped them achieve their goals. On this premise, it could be argued that the participants' deliberate usage of these devices has turned them into significant constituents of their translation practice in Wikipedia. Moreover, even though none of the devices listed above is essential for the performance of the practice itself, they are very often constituents of it. For example, materials such as dictionaries, which serve multiple purposes, have long been works of reference for translation practitioners (Olohan 2021). Others such as databases and MT have become increasingly important over the years. Yet, as 6.2 and 6.3 will examine, the materials more directly involved in the development of translation in Wikipedia are likely to be found within the confines of the user-driven encyclopaedia.

## 6.2 Usage of content-creation devices in Wikipedia: Bots and CX

This section of the analysis tackles the participants' views on, and use of, bespoke cross-lingual content creation devices in Wikipedia to better understand how such devices may have configured their practices. As stated in 1.1.2 and 2.4.1, the term 'content creation devices' encompasses both bots and CX. Furthermore, as it was argued in 1.1 and throughout this thesis, it is difficult and impractical to disentangle translation from other forms of editing in Wikipedia because they both form a continuum (Shuttleworth 2017; Jones 2018b). Consequently, the following paragraphs will address bots and CX as devices that have been mobilised differently by the four Wikipedia communities with the specific purpose of creating multilingual content. These differences in mobilisation may have contributed to configuring the participants' practices in a distinct manner. To ascertain whether

this is the case, the analysis will draw on data elicited during a specific block of the interviews that focused on the use of automated devices.

The first subsection will concentrate on bots. Upon analysing the interviewees' views on these devices, 6.2.2 will shift the focus to their attitudes towards CX. Each subsection will analyse the most relevant themes that surfaced during the interviews. Finally, the impact of both devices on the evolution of the participants' practices will be examined in 6.3.

### 6.2.1 Bot-generated content

As discussed in 1.1.2, to date, research on bot usage in Wikipedia remains scarce. Despite the fact that these devices are almost as old as Wikipedia itself, studies on bots have not aroused much interest in academic circles. Except for a few notable exceptions (Geiger and Halfaker 2017; Tsvetkova et al. 2017), most investigations into the deployment of bots seem to have occurred in the early 2010s. Notwithstanding this datedness, the research undertaken by Geiger (2011) and Halfaker and Riedl (2012) provides some valuable insight into how bots, albeit not bot creations, were regarded at the time. For example, while some editors would view them as potentially disruptive (Halfaker and Riedl 2012, 81) and 'ruthlessly moral' (Geiger 2011, 91), perceptions around their usage in Wikipedia were generally positive (Geiger 2011; Halfaker and Riedl 2012).

In Chapter 4, when examining the negotiation process behind the approval of content-creation policies, it was observed that views on bot-generated material in the Spanish Wikipedia were varied and at times irreconcilable. Although the data were gathered from a debate that unfolded in 2008, contemporary Dutch and Swedish Wikipedia editors (see 4.3.3 and 4.3.4) indicated that disagreement over bot-produced Wikipedia entries was still a recurrent phenomenon in their communities as of April 2020. To gain more insight into the matter, part of the interviews conducted between July and August 2020 addressed the participants' views on bot-produced Wikipedia entries within the broader framework of multilingual content creation. Table 6-2 below shows the most salient themes that emerged during the discussion.

| Themes | | | |
|---|---|---|---|
| **Arguments in favour of bot usage** | **Participant(s)** | **Arguments against bot usage** | **Participant(s)** |
| As a starting point | Alex (ES), Jos (NL), Leslie (NL), Nik (NL), Alva (SV), Charlie (SV) | Poor quality of content | Maxime (FR), Sam (FR), Guus (NL), Leslie (NL), Nik (NL), Alva (SV), Charlie (SV), Robin (SV) |
| With human supervision | Ariel (ES), Pau (ES), Dominique (FR) | Bot overuse can be damaging | Ange (FR), Maxime (FR), Guus (NL) |
| To fill knowledge gaps | Kim (SV), Robin (SV) | Improvement in bot performance required | Charlie (SV), Kim (SV), Robin (SV) |
| AI has improved | Pau (ES) | Bot-generated articles outdated | Guus (NL), Kim (SV) |
| Good for smaller Wikipedias | Cris (ES) | - | - |

Table 6-2. Themes and subthemes extracted from the participants' views on bot-generated content.

As Table 6-2 illustrates, views on bot-generated content in Wikipedia are still mixed across the four language communities. Nevertheless, unlike the opinions expressed by Spanish Wikipedia editors in 4.3.1, the interviewees' points of view were slightly less rigid. For instance, six participants – Leslie (NL), Nik (NL), Alva (SV), Charlie (SV), Kim (SV) and Robin (SV) – offered arguments for and against bot deployment. Another aspect worth highlighting is that, despite the lack of a substantial number of bot creations in the Spanish Wikipedia, all four participants were in principle in favour of using automation. These views contrast with the

arguments given by three of the French Wikipedia interviewees, who approached the use of bots in more condemnatory terms.

Regarding the two Wikipedias known for their reliance on bots, the opinions expressed by the participants were generally aligned with those that formed part of their communities' shared repertoire. Thus, with the notable exception of Guus, the Dutch Wikipedia participants were welcoming of the use of bots aimed at the creation of encyclopaedic content. All four Swedish Wikipedia counterparts were overall supportive of bot usage, but their approval was contingent on the bots being enhanced. The next paragraphs will provide more context to all the participants' responses. In doing so, the analysis will show that the interviewees' arguments for and against were at times less nuanced than the data representation in Table 6-2 may suggest.

### 6.2.2 Participants' views on the advantages of mobilising bots

Six participants regarded bots as optimal devices to initiate Wikipedia articles on topics that, due to their specificity, are unlikely to be created by human editors. As stated in 2.4.1, such articles include but are not limited to small towns, landmarks, and a wide range of animal and plant species. Of note is the fact that none of the interviewees identified themselves as bot users, and instead, their views were based on their perceptions as engaged members of their communities. Dutch Wikipedia editor Jos, for instance, opined that:

> Bot-created articles are great if you need some kind of groundwork or some kind of infrastructure that you plan to build on top of. Not for some species of beetle that is rare, that is barely found in the Amazon rainforest. Like, there is nobody in the Dutch language [Wikipedia] are that is actually going to write about that.

Their colleague Leslie expressed a similar opinion, adding that bot-generated content does not 'weaken' Wikipedia. To them, these automated devices are an asset because 'bots can create very basic articles that someone else can expand later, following the principle of collaboration that characterises Wikipedia'. The participant also objected to changing their community's policies on bot usage, noting that although some bot-generated entries may contain little information, 'an existing [Wikipedia] article will always increase the prospects of someone else

stepping in and adding more data'. This concept of expansion also emerged during the interview with Nik, for whom bot-generated entries 'can help newcomers to start editing and adding content because the framework is already there'.

Concerning the Swedish Wikipedia, attitudes to bot usage were generally positive, but the participants' answers leaned towards caution. According to Alva, some balance is necessary after the disruption caused by Lsjbot, the automated device behind 80% of the Swedish Wikipedia entries ('Wikipedia Statistics' 2019). This concern was shared by Charlie, who stated that at the beginning, they were satisfied with Lsjbot's performance, but their opinion changed when:

> Lsjbot moved on to geography, and it became a catastrophe. It had... The references were not as good, so there were many mistakes in the articles, and there were doubles, where the bot wrote two articles, or three, about the same place.

Regardless of the problems stemming from bot usage, Charlie insisted that they were still 'grateful for the articles on biology', many of which would 'have never been created without Lsjbot'.

Perhaps to prevent errors such as the ones described by Charlie, another three participants declared their support for bot deployment subject to more stringent policies and rigorous human supervision. For Pau (ES), who believes that AI 'is much better than a decade ago', bots can contribute to enriching the content available in the Spanish Wikipedia. To achieve this, the participant advocated a major integration of devices to the editors' practices, starting with the gradual automation of mundane tasks such as adding references, categories and images to the text. Pau went on to add that for this mobilisation of devices to prosper, 'a human must review what the bot does'. Likewise, Ariel (ES) and Dominique (FR) observed that, if more robust policies were in place, bots could be used to make knowledge more accessible to non-English-speaking readers.

This accessibility is what may have prompted Swedish Wikipedia editors to endorse bot usage almost a decade ago. According to Kim and Robin, bots fulfil a crucial function in so far their deployment is intended to fill knowledge gaps. When asked to elaborate on their answer, Robin remarked that 'bots can obviously help to create articles in a quicker way, and help to cover entire areas of knowledge so that there's at least a foundation to work from'. Likewise, Kim stated that although bots were

not conceived as translation devices (see 1.1.2), at the beginning they were received 'quite favourably' as cross-lingual devices by the Swedish Wikipedia community because 'the articles hadn't been there'. For the participant, bots were initially welcome because even though '[bot-generated] articles were quite stereotypical and had short sentences with quite basic variation or not any variation at all, they told the truth and were well-sourced for the most part'.

Finally, for Cris (ES), the use of bots as content-creation devices boils down to the number of speakers in a given language. The participant opined that the mobilisation of bots in the Spanish Wikipedia as integral components of editing practices was neither justified nor required because the language is widely spoken. Nevertheless, Cris viewed bot deployment with more positive eyes in smaller communities such as the Swedish Wikipedia because they only have 'a few active users and, in the case of the Swedish version, the language is spoken only in Sweden [and Finland]'. Despite the benefits, this use of bots has had its downsides. The next subsection will address the participants' views on the negative impact such devices have had on their communities' practices.

### 6.2.3 Participants' views on the negative impact of bots

As was explained in 6.2.1, some participants offered insight into both the advantages and the drawbacks of bot deployment on content creation practices. For eight interviewees, the quality of bot-produced Wikipedia entries was significantly inferior compared to those written without the device. Considering that the Swedish Wikipedia has the highest percentage (81%) of bot-generated entries of all four Wikipedia communities, it comes as no surprise that some of the major objections to the mobilisation of these automated devices were reported by editors of that community.

According to Alva (SV), 'there has been a lot of discussion about the quality of the [bot-generated] articles' in their community, and much of what was created by Lsjbot is now 'under review'. Along the same lines, Charlie (SV) argued that, despite Lsjbot ceasing its functions in 2016, 'there's still a lot of crap [in the Swedish Wikipedia], and there's a lot of cleaning needed'. The participant went on to give a more detailed account of how the widespread mobilisation of the device

up to that time has resulted in a long-term burden for Swedish Wikipedia editors, who have had to devote more time to 'cleaning' the site:

> So, the Swedish Wikipedia got bigger and bigger, more and more articles every week, when the bot was going. Since two and a half years, Swedish Wikipedia still has a lot of writers, but with all the cleaning, the Swedish Wikipedia doesn't get more articles now, it gets less, because so many articles are deleted for their bad quality. So, all the articles we write every week, every month, don't make Swedish Wikipedia bigger.

On the same issue, Robin (SV) regarded some Lsjbot-generated articles as being of 'a very low quality'. In their view, the Swedish Wikipedia community has not necessarily changed its attitude towards bots. Instead, a decision was made to stop producing bot-generated content because the editors could no longer 'handle them'. Charlie noted that the decision to discontinue Lsjbot did not sit well with the developer of the device, as noted in 1.1.2:

> There were voices saying, 'you have to stop so we can discuss the quality before we get more articles', and he stopped the bot and got a lot of criticism, so he stopped making bot articles for Swedish Wikipedia. I was very neutral during that discussion, but it ended with the Lsjbot maker getting very... He wasn't happy at all; he thought that Swedish Wikipedia should have been more grateful for all the articles he had made.

Kim (SV) remarked that part of the community's discontent with Lsjbot stemmed from the unreliable, and sometimes outdated, sources the device relied upon. The interviewee asserted that 'at least a million of those articles are still problematic' and added that, as they stand, some of them 'do not belong in Wikipedia'. Charlie and Robin expressed similar opinions on the need for improvement, with the latter remaining open to a potential deployment of bots in the future if the device were to gather data from valid references. For Robin, the bot debate goes beyond whether using them is 'good or bad'.

The situation in the Dutch Wikipedia, another community known for its past dependence on bots, shows some parallelisms with the case described above. However, as observed in 1.1.2, the percentage of bot-produced entries there, though high, is significantly lower (51%). Guus, the only participant of the Dutch Wikipedia to oppose the use of the device, said that bot-generated content was outdated. When asked why they thought no one had modified the information, they responded:

> These [animal and plant] species change all the time, they are changing, due to genetic research or so the names will change, the families they are in will change, so it's just a moment in time they uploaded these bot articles, but I don't think they are correct anymore. And no, nobody is bothering about it; I think the people who did it may be gone, I don't know. So, it's... Yeah, there might be many errors or not up to current standards.

The statement above suggests that extensive bot usage led to a similar outcome in both the Swedish and Dutch Wikipedia communities. Notwithstanding these similarities, the answer Guus provided implies that their fellow Wikipedia editors eventually gave up on updating the information. To put it more concretely, the Dutch Wikipedia community seems to have carried on editing without paying much attention to those bot-generated entries. This action differs from the one taken by the Swedish Wikipedia, which, as the participants expressed, seems to be still committed to amending the articles created by Lsjbot.

Nik noted that standards of quality in the Dutch Wikipedia have changed over the years. According to the participant, the tendency nowadays is to edit without bots because there is an expectation that every new Wikipedia article should contain 'at least a couple hundred words'. Nik's answer indicates that content-creation practices in their community have evolved to become stricter. These changes to the shared repertoire mean that devices such as bots have lost their value because the type of tasks they used to perform is no longer required. In other words, the inability of bots to meet the practitioners' demands led to their discontinuation. Leslie also expressed a similar view, stating that bot-produced Wikipedia entries resemble 'database stuff that has more its place in Wikidata[42] or other platforms'. Simply put, the participant believed that articles created by bots lack meaningful content, and they tend to include basic information, primarily figures. Guus opined that if someone were to deploy bots these days, their actions 'would face criticism'. This shift in the standards is marked by a process of dissociation whereby bots went from being considered critical devices for the production of multilingual content to becoming an inconvenience.

Such criticism also comes from other Wikipedia communities that have traditionally opted to restrict bot usage. For example, Ange (FR) referred to the

[42] As explained in 2.4.1, Wikidata is an online repository that serves Wikipedia and other sister projects.

situation in the Swedish and Cebuano Wikipedia communities as ‘insane’, adding that having deployed such ‘number of automatic translation bots’ was not ‘a good idea’. Likewise, Maxime (FR) commented that while in the ‘Swedish and Cebuano Wikipedias you have a lot of content, that’s just empty content’. Regarding the possibility of implementing the device in the French Wikipedia, Maxime observed that their community members – positioned as brokers (Wenger 1998) – learned from the errors others made. As examined in 4.2.1, the French Wikipedia has quality as an essential value. Therefore, in Maxime’s view, the type of articles that bots can generate is incompatible with the values upheld by their community. Similar to what Leslie said, Maxime commented that bots import ‘stuff that has been picked from Wikidata, which doesn’t contain much information’.

Thus far in the analysis, it is apparent that despite their perceived advantages, bots have fallen out of use as cross-lingual content-creation devices and consequently as configuring elements of practice. The reasons that prompted the Dutch and Swedish Wikipedia communities to bring their deployment to a halt seem to be rooted in a more rigorous application of editing policies such as verifiability and notability. Furthermore, as values such as quality evolved over the past years, the mobilisation of bots became incompatible with the new framework in place. Because of their larger and more diverse number of speakers, other communities such as the Spanish and French Wikipedias seemed to have thrived without bots. In the case of the French Wikipedia, the scepticism expressed by a few participants over the implementation of such devices is also guided by their observation of the events that transpired in the Swedish and Cebuano Wikipedia communities.

### 6.2.4 The impact of CX in configuring the translators’ practices

As stated in 1.1.2, since its inception in 2014, CX has gained widespread popularity as a configuring element of practice across various language communities of Wikipedia. The impact of this device on the editors’ practices was first documented by McDonough Dolmaya (2017), who situated the number of CX-generated articles at 94,210. In the four years that have elapsed since her investigation, the number of Wikipedia articles created with CX has grown almost ten-fold (‘Content Translation Statistics’ 2021). Such growth has been more pronounced in the

Spanish and French Wikipedia communities, which rank first and second in the number of CX-generated articles. To better understand the reasons behind the increasing usage of this device and its impact on translation in the user-driven encyclopaedia, the 16 participants were asked the following questions: 'Have you used Wikipedia's Content Translation Tool (CX)?' 'If so, could you tell me more about your experience using the tool?' 'What were the pros and cons of using CX?'

Except for Alex (ES) and Charlie (SV), all the participants declared having incorporated CX into their practice. Of these, 12 identified themselves as users of the device and two as partial users. As regards language community, the data show that the mobilisation of CX was less common among the Swedish Wikipedia participants. Both Alva (SV) and Kim (SV) indicated that their usage of CX was peripheral, while Robin (SV) commented that for them the device was only useful as a starting point, 'not as the finish line'. This difference in usage tallies with the low mobilisation of CX in the Swedish Wikipedia, especially when compared to other communities of a similar size. At the time of writing, the Swedish Wikipedia remains in the 49$^{th}$ position with 3,045 CX-generated articles, well below smaller communities such as the Asturian, Basque, Galician and Norwegian (both Bokmål and Nynorsk) Wikipedias ('Content Translation Statistics' 2021).

On the advantages of deploying CX, most participants agreed that the device had contributed to optimising their time. Thus, a common theme that surfaced during the interviews was that CX had made translation easier and faster for the editors. According to Ariel (ES), CX 'helped a lot, it makes it easier for you, because there used to be other tools where translations were quite clumsy'. Cris (ES), who said they had been using CX since they were a beginner, opined that the device 'is good when it's necessary to do some quick translation'. In a similar vein, Pau (ES) observed that in the past, CX had been 'a big help' to them and that 'the idea and initial dynamic of the tool were good'.

Along the same lines, Ange (FR) noted that while CX was far from perfect, they were 'quite happy with it' because:

> All the sections are aligned because sometimes it's not the same length when you translate from one language to another, so it's good to have them always aligned. This alone is good, and when you put templates or references, it kind of gets most of the time how it works and translates it

> well too, so that's nice also. Sometimes you have to check and correct it sometime after, but most of the time it works, so it's good.

Dominique (FR) and Maxime (FR) expressed similar opinions. For Dominique, CX 'really helps to translate much faster'. Their colleague Maxime provided a more detailed account by adding that CX had 'made the translation experience much easier' partly because 'nowadays you have a tool that takes care of all the technical issues and all the conversion systems you had to deal with before'. Equally appreciative of the affordances of CX, Sam (FR) commented that the device had configured their practice:

> I would say [I use CX] most of the time now. At the beginning, I wasn't [a user of CX], and it was taking two to five times more time for me to translate the page. Now, I think, also the automatic translator has improved. And what is good is that on the side it says how much of each paragraph comes from the original and has not been corrected, so it gives me a way to control what I'm doing.

Although the Dutch Wikipedia currently occupies the 28th position in terms of CX-translated articles, the participants' answers suggest that usage of the device may be gaining ground in the community. For instance, Guus, who had previously been critical of bot-generated Wikipedia articles, noticed that CX 'is a good tool in general because it takes everything you need: the internal links and so are correct, the categories…so it's very easy [to use]'. Jos observed that since the device was first launched, there have been 'incremental changes'. According to Leslie, such improvements in automation prompted them to incorporate CX to their practice: 'I did do translation, but much less, because... Yeah, it was less easy, so the tool really changed for me the balance of it'. The impact of CX in configuring the practice was also attested by Nik:

> Yes, it makes me... A lot easier to translate. I've done [it] by hand in the past, more than five years ago, and I copied the original source of the English Wikipedia and pasted it in a word processor to document my work. Then I would translate sentence by sentence. It's very hard with all the templates and references. And the content translation [CX] does a very good job in matching templates from the English and Dutch Wikipedia.

As explained at the beginning of this section, the mobilisation of CX among the Swedish participants was limited. Nevertheless, those who had resorted to the device were satisfied with its overall performance. Alva, for example, noticed that

CX had saved them time because they ‘don’t like to start with an empty page’. Kim was of the view that ‘automatic translation tools [such as CX] are useful aids, but are not to guide you and take over the responsibility that you, as an editor, have to present things to the world’. Like most other participants, Robin also pointed out that ‘the advantage of using the translation tool [CX] is that it speeds up – it makes the process quicker, whereas there aren’t many advantages to doing it manually’.

Despite the benefits, the participants’ answers indicate that deploying devices such as CX often comes at a cost. For instance, six participants mentioned that CX was not always efficient for translating wikicode such as templates, bibliographic references and infoboxes. Another four interviewees reported that CX sometimes failed to provide the correct translation of specific lexical items. A smaller number of participants complained about problems with the interface, the requirement to be registered in Wikipedia to run CX, and the lack of diversity resulting from a surfeit of translations from English.

Regarding difficulties with wikicode, three of the six participants that reported bugs with CX were from the Spanish Wikipedia, the community that features as the top user of the device. The other three participants were from the Dutch Wikipedia. According to Ariel (ES), CX ‘tends to modify templates’ even if it is not necessary. Cris (ES) stated that the device ‘doesn’t recognise which templates aren’t available in Spanish and English, and it doesn’t know how to recognise categories’. Pau (ES) remarked that since CX is not able to translate all wikicode, editors have to spend more time post-editing. The participant added that sometimes they experienced ‘frustration’ because CX was not capable of translating basic headings and sections of an article such as ‘references’ and ‘external links’. Guus (NL) noted that ‘infoboxes are not translated sometimes because [the source and target Wikipedias] don’t have the same parameters’. Jos (NL) found this mismatch between language communities ‘super annoying’ because ‘they take a lot of time to fix’. Likewise, Nik (NL) argued that those persistent bugs were a hindrance because they forced editors to ‘close [the parameter] and then open it through the translated box and fill it in manually’.

Another major drawback reported by the participants was the inability of CX to provide accurate translations. Ange (FR) stated that CX ‘gets lost and confused with

very specific terminology', hence posing additional challenges to editors. Sam (FR) commented that, similar to other automated devices, CX does not always retrieve meaning in context. The interviewee added that to tackle these issues, it was 'necessary to go paragraph by paragraph and redo the translation in some parts'. Guus (NL) pointed out that 'although [CX] has improved a lot, it still fails to translate words in the correct order'. Kim (SV) contended that a careless use of CX could lead to the proliferation of false friends. The Swedish Wikipedia participant was critical of what they perceived as a 'dependence on technology' in translation.

The third reported issue with CX was related to the interface. Ariel (ES) remarked that the device was 'not visually attractive' and that the screen tends to narrow down the text to the extent that sometimes they accidentally 'skip a line or two' during the translation process. Moreover, the participant said that once someone starts a translation on CX, no one else can work on the text until it is published in Wikipedia. Maxime (FR) raised a similar complaint, noticing that when using CX, 'it's not possible to translate an existing article or to publish part of the translation'. According to Leslie (NL), another limitation of the device is that 'at times, you will click on a paragraph [to translate] and the tool will just tell you that it failed to load'. Dominique (FR) recounted bugs with algorithms when CX was at an early stage of its development: 'I've been given a suggestion to translate articles on American footballers, which I'm sure are very interesting, but they're really not my field of interest'.

Finally, Alva (SV) considered that the requirement to be registered in Wikipedia to access CX could dissuade potential editors from contributing to the project. Kim (SV) argued that CX was impractical for them because they preferred to retrieve information from more than one language version of Wikipedia. Robin (SV) said that they were against using CX on a regular basis because the device 'doesn't create much diversity between language versions' of Wikipedia. In other words, the participant contended that a major reliance on automated devices such as CX leaves editors with little room for manoeuvre. Cris (ES) expressed a similar opinion when they explained why, in their view, CX was so popular in the Spanish Wikipedia:

> The English Wikipedia has always been seen as sort of an ideal to strive for. Most of my colleagues who have good English skills prefer to read articles on the English Wikipedia instead

> of the Spanish one because they think it's more up-to-date, or more accurate, or respects neutrality more in some situations…

As stated in 1.1.1, previous research by McDonough Dolmaya (2017) has addressed concerns about the lack of diversity that could result from an overreliance on the English Wikipedia as a primary source for translations. The data analysed in this section have shown that, regardless of such considerations, the vast majority of participants incorporated CX into their practice. Moreover, the participants' reports on their experience running the device indicate that, for better or for worse, their performance was configured by it. In most cases, the inception of CX provided the participants with a shortcut to slower manual translations. This quicker route, however, presented a few obstacles that the interviewees had to overcome. For instance, glitches and technical limitations forced some participants to devote more time to proofreading and post-editing. In the next section, the impact of automation on translation is further examined through the participants' perceptions of how the practice has evolved in Wikipedia.

## 6.3 The evolution of translation practices in the four Wikipedia communities

Towards the end of the interview, participants were asked whether they had perceived significant changes in how people translate in their Wikipedia communities over the last years and, if so, whether they thought automatic tools had contributed to those changes. Participants were also encouraged to comment on how they envisaged the future of translation in Wikipedia and the role of Wikidata in that scenario.

Half of the interviewees believed that machine translation had improved significantly since they first joined Wikipedia. Another five participants envisaged further changes to translation in the coming years with the launch of a new cross-lingual content-creation initiative called Abstract Wikipedia. The same number of participants opined that Wikipedia had become stricter regarding automatic translations. Two participants argued that the mobilisation of new devices has often been met with resistance from a reduced group of conservative senior editors. Table

6-3 below shows these and other themes that emerged during the last phase of the interview.

| Theme | Participant(s) |
|---|---|
| New advances in automation | Pau (ES), Ange (FR), Dominique (FR), Sam (FR), Leslie (NL), Alva (SV), Charlie (SV), Kim (SV) |
| Abstract Wikipedia | Cris (ES), Pau (ES), Ange (FR), Maxime (FR), Jos (NL) |
| Wikipedia has become stricter | Cris (ES), Guus (NL), Alva (SV), Charlie (SV), Robin (SV) |
| Some senior editors resist change | Alex (ES), Ariel (ES), Maxime (FR), Jos (NL) |
| Google outpaced Wikipedia | Nik (NL) |

Table 6-3. List of changes in Wikipedia translation resulting primarily from the mobilisation of automated devices.

As regards advances in automation, Pau (ES) noticed that since they joined Wikipedia back in 2009, 'quite a few apps have appeared, and we have seen an improvement in the way translation is done overall'. When asked to elaborate on their response, Pau added that 'the content translation tool [CX] makes a big difference; it was the right thing to do'. For Ange (FR), external devices such as Google Translate have also contributed to configuring how translation is performed in Wikipedia: 'Google Translate improved, also, over the years; probably that helped'. Dominique (FR) stated that 'we [in Wikipedia] already have something that is much more impressive than what we had a few years ago'. Their colleague Sam commented that since CX was first launched in 2014, 'more people have joined the Open Source movement'.

Leslie (NL) observed that while translation practices in Wikipedia have evolved as a result of automation, devices such as CX have also undergone significant changes. According to the participant, these changes stem from the interaction between materials and practitioners: 'We're sort of feeding this tool with the choices we

make, maybe it can learn something from it'. Alva (SV), who created their Wikipedia account in 2009, noted that 'automatic translations are gradually improving, so the tools are going to improve, and there will be more people using machine translation'. Similarly, Charlie (SV), in Wikipedia since 2012, declared that automation has evolved throughout the last five years. The interviewee compared the evolution of translation devices to that of chess programs: 'It's like chess: from the beginning, chess programs were very bad, and you could beat them, but nowadays nobody can beat a chess program if it's good'. Kim (SV), who registered in 2006, argued that although CX has come a long way since it was first launched, the device 'can't give us its full potential yet'. However, they were hopeful about the future: 'maybe in three or four years' time, the translation tool [CX], like all similar translation tools, can be more of a practical colleague in our article creation'.

As previously stated, five interviewees envisaged significant changes in the translation landscape with the ongoing development of Abstract Wikipedia and its expected launch between 2022 and 2023. Because the project is still at its earliest stage of development, there is not much information available. However, a page on Meta-Wiki, where Wikimedia projects are coordinated, states that the concept of Abstract Wikipedia was devised 'to let more people share more knowledge in more languages' ('Abstract Wikipedia' 2021). While this phrasing may come across as ambiguous, the next paragraph narrows down the project aims:

> Abstract Wikipedia is a conceptual extension of Wikidata. In Abstract Wikipedia, people can create and maintain Wikipedia articles in a language-independent way. A Wikipedia in a language can translate this language-independent article into its language. Code does the translation.

According to the page, 'language-independent' means that, rather than being created in a specific language, articles will be generated by combining abstract codes or functions from a new Wikimedia project called 'Wikifunctions' and material gathered from Wikidata. The combination of both functions and metadata will then be fed into different language versions of Wikipedia. At the time of writing, there are no examples of what a Wikipedia article created this way would look like. Nevertheless, it is apparent that the Abstract Wikipedia initiative seeks to bridge the existing gap between language communities of the user-driven encyclopaedia

and give everyone access to free knowledge in their language, thus achieving Jimmy Wales' 'prime objective'. If the project comes to fruition, it seems likely that Wikipedia may no longer rely on editors to perform the translations. The materialisation of this project would lead to significant changes, especially when considering that, although certain communities of the encyclopaedia have a history of mobilising bots to fill knowledge gaps, so far, the practice of translation has remained a human endeavour.

Cris (ES) referred to Abstract Wikipedia as a major turning point in the documented history of translation in the user-driven encyclopaedia. The participant said that it would be 'a new gateway into translating conventional articles'. Maxime (FR) hoped that by gathering information from Wikidata, Abstract Wikipedia would help 'solve the problem we currently have with templates and CX'. Pau (ES) and Ange (FR) regarded the launch of the project as a step towards a consolidated integration, breaking the barriers between language communities. On the other end, Jos (NL) expressed concerns about the negative impact that Abstract Wikipedia could have on smaller Wikipedia communities:

> While it sounds great to have a core set of a few thousand articles available in whatever language you want, at the same time, it's a huge danger for community development. I'm sure you're familiar with the literature about, like, what amount of seeding helps a community and when it starts hurting a community. If you're seeding a community with too much content, there is an actual risk that the community will never develop. And especially if you're talking about smaller languages, that's a pretty big danger.

While resources such as Abstract Wikipedia may influence translation in the years ahead, five participants posited that values and expectations around the practice and the materials that configure it have already changed in their language communities' shared repertoires. For instance, Cris (ES) contended that:

> If someone made an article like that [with little information] and published it today, it wouldn't last five minutes. I think that in all these years, since Spanish Wikipedia started until today in 2020, it has become stricter regarding the quality that the content should have, and I say this as a good thing.

Likewise, Guus (NL) noted that, when they started contributing to Wikipedia in 2002, the encyclopaedia 'was empty, so [they] wrote a lot of articles and it was very easy; [they] could write from heart, from what [they] knew about, but this has

changed, because now it's not allowed, and people have to use sources'. In a similar vein, Alva (SV) observed that 'policies [in the Swedish Wikipedia] have become stricter when it comes to sources and references'. Charlie (SV) believed that this tightening of policies increased the levels of trust in the quality of the content available in Wikipedia:

> You can read it, you can trust what's there, and if you can't trust it, there's often a template at the top that says 'no references', or 'better references needed', or 'this is a person's point of view', or something like that.

Robin (SV) argued that more efficient quality controls have contributed to a major integration of translation and editing practices in the Swedish Wikipedia:

> In the early days of Swedish Language Wikipedia, a lot of the featured articles were translations of featured articles in English or German or French or other languages, but it's not considered possible for an article that is only translated to be featured or to reach a prominent status, so you would have to go through all the sources yourself or create an article from scratch yourself; so there has been a change in the perception of the quality of translated articles.

As mentioned throughout the thesis, the strict enforcement of Wikipedia policies is often the responsibility of senior editors. Due to their long-term mutual engagement and under the belief that what works well does not need to be altered, senior editors may resist innovation (Wenger 1998). In various sections of Chapter 4, it became apparent that a few editors resisted changes to automation and specific aspects of the documented standards, leading to lengthy discussions. This issue came to the surface again during the interviews, with four participants stating that resistance to change was common among some editors of their Wikipedia communities. One illustrative example of such resistance is found in the account given by Alex (ES), who criticised what they perceived to be an increasing over-reliance on automated devices as constituting elements of everyday practices. The participant, registered since 2007 and one of two non-users of CX, noted:

> Many people, in the last years, have turned to being bad readers and bad writers. They just type short messages in their smartphones and that's the most things they just write. And they maybe trust too much those clumsy automatic tools. Hear me how I talk. Those *clumsy* automatic tools, yeah?

The remaining three participants were not reluctant to deploy automation, but they had experienced resistance from others in their Wikipedia communities. Ariel (ES),

in Wikipedia since 2015, noted that senior editors in their community had long opposed the translation of navigation templates from English into Spanish:

> I've been trying to create more of these recently. When I first started I did... I wanted to do them, especially because in English they were used a lot, but old users removed them; they deleted the templates. So now instead I try when I have already finished all the articles related to the information I have... I think I have only around three or four right now. One that is for the human microbiome, the structure of the Earth, one on chemical equilibrium, and distillation. I always make them with a little bit of caution; as I was saying, older users don't like these templates.

Therefore, the participant had to adapt aspects of their practice as part of their engagement with fellow editors (Wenger 1998; Warde 2005). In particular, Ariel had to moderate their inclusion of devices such as navigation templates, which are allowed in the English Wikipedia, to conform to the expectations and values that are part of the Spanish Wikipedia shared repertoire.

Maxime (FR) observed that when the Visual Editor was first incorporated to Wikipedia between 2012 and 2013, the mobilisation of the device faced opposition in the French language community:

> Yeah, the content translation [CX] is built on the Visual Editor. And I personally only use the Visual Editor; it's much easier, simpler, don't have to think about messing up the code. And, again, you have the same issues: people are against it because 'it's not the right way to do it', 'it's too simple', 'it's not covering everything' even if it's covering 97% of everything. We have lots of conservative people in our communities, that's a problem. So, when you try to bring a new tool, they're always like, 'are you sure it's the right way to do this?' or they're like 'you're trying to change the Wikipedia I know'. Yeah, that kind of topic I can discuss.

Similarly, Jos (NL) argued that some editors in the Dutch Wikipedia were opposed to the deployment to automated devices regardless of their benefits:

> I hope that the attitude will change a bit towards accepting content. I see how people don't like automated translation as-is; like, you don't want people to dump automated translations, but I think people should loosen up a little and get a life. Like, just focus on things that are more important; if the overall quality is fine then great, but I don't hope that people just start mass translating articles because I think that's bad.

Finally, Nik (NL) reported that a relatively recent development concerns how 'Google outpaced Wikipedia' regarding automated translations. Similar to what Abstract Wikipedia is projected to achieve, Google retrieves information from

Wikidata and generates an automatic translation of the source. Although this process is external to Wikipedia, it is elucidative of how the user-driven encyclopaedia, which provides the infrastructure for the performance and configuration of several practices, can also serve as a device for those who seek first-hand information. After all, as discussed in 2.4, materials can, and often do, adopt different roles depending on the practice in which they are mobilised (Hui 2017; Shove 2017; Olohan 2021). The following subsection on Wikidata offers an illustrative example.

### 6.3.1 Wikidata as a device and its impact on Wikipedia translation

As explained in 2.4.1, Wikidata is a free and open knowledge multilingual collaborative repository first launched in 2012. At the time of writing, the platform contains ca. 95 million items. According to information retrieved from Wikidata's front page, the site 'acts as a central storage for the structured data of its Wikimedia sister projects, including Wikipedia' ('Wikidata Main Page' 2021). Therefore, since its inception almost a decade ago, Wikidata has served as a repository for all language communities of the user-driven encyclopaedia. The fact that structured data or metadata such as numbers, figures and personal information about notable people are now contained in one digital repository instead of being stored independently by individual Wikipedia communities has ensured more cross-lingual consistency and accuracy.

Wikipedia presents the data gathered from Wikidata mostly through infoboxes but also via categories and links. To put it more concretely, most Wikipedia articles are linked to a single Wikidata item containing data about them. Another reported advantage of this centralisation of knowledge, apart from coherence, is that when a community publishes a new Wikipedia article linking to Wikidata, that article is automatically connected to all the language communities where the content is available. This process has been integrated into content-creation devices such as CX, which generates the list of languages (interwikis) once the article has been first posted on Wikipedia.

Considering that Wikidata has been incorporated into translation, it can therefore be argued that the repository is a device that has contributed to configuring the practice. Besides its integration into CX, Wikidata's multilingual philosophy promotes the translation of content into other languages by showing what information is missing. Moreover, as is the case with Wikipedia, a large amount of the structural data stored in Wikidata is only available in English. Therefore, the metadata shown in infoboxes, unless translated, will be displayed in English regardless of the language version of Wikipedia in which they are used. To prevent these instances of language mixing, Wikipedia translators concerned about these issues may devote additional time to rendering the information into the target language by editing the source directly in the Wikidata item. For the sake of illustration, Figure 6-1 below shows the Spanish Wikipedia article on the Russian physician Lev Levin, translated from the English Wikipedia. The infobox template imported from Wikidata, on the right, displays information in Spanish about the subject's birth, death, cause of death, and nationality. However, the information on Levin's education – as indicated by the red arrow – contains the name of the Russian institution where he studied, but this is given in English ('Lev Levin' 2021).

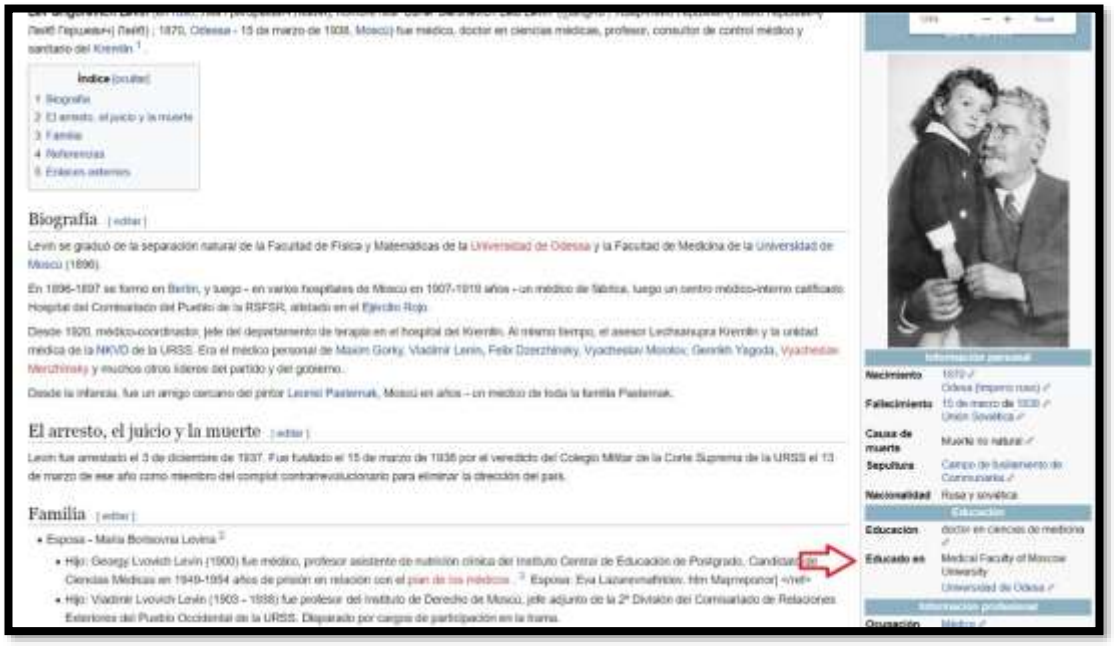


Figure 6-1. Spanish Wikipedia article on the Russian physician Lev Levin, translated from the English Wikipedia.

In light of this language mixing, the last part of the interview was aimed at ascertaining the potential impact of Wikidata in configuring translation practices in

Wikipedia. As discussed at the beginning of 6.3, participants were asked to describe the role of Wikidata in their practice as translators, particularly regarding infoboxes. Fourteen participants answered the question in some capacity, while only two – Nik (NL) and Kim (SV) – declared that they were not familiar with Wikidata and its influence on Wikipedia. Those who addressed the subject were of the view that the repository had been a positive contribution to the encyclopaedia. However, a few argued that certain elements such as templates required further attention.

Alex (ES) stated that they had encountered no problems with Wikidata. However, the participant was of the opinion that 'content translation must be kept separately, as an autonomous area, not tied to algorithms'. Ariel (ES) commented that they had resorted to Wikidata exclusively for the so-called 'authority control', a process through which a Wikipedia article is assigned a unique identifier. The authority control is usually found at the bottom of each Wikipedia article that has been linked to the Wikidata item. Ariel recognised that they were not conversant with the most advanced features of Wikidata and that their usage of the repository had been restricted to linking the translated article to other language versions of Wikipedia. Cris (ES) said that Wikidata is 'a project [they] love' and that at the time of the interview they had put their activities in the Spanish Wikipedia on hold to invest more time in expanding the repository. The interviewee referred to the impact on the Wikipedia infoboxes as 'generally quite positive', although they remarked that there was 'still a big difference between what is available in English and Spanish on Wikidata'. Likewise, Pau (ES) mentioned that Wikidata had contributed to configuring their practice by relieving them of having to perform tedious tasks:

> These are the kinds of things that help [with Wikidata]: you do that, you improve the template, and if this template is used in ten thousand articles, suddenly, after you edit the code to import that data from Wikidata, suddenly ten thousand articles where so far there was nothing there get the data. I mean, that is huge. I did something also in the past, to be honest, I spent... Maybe one year or one year and a half ago I improved some templates, but I felt very alone.

Ange (FR) regarded the repository as a 'game-changer and a time-saver', and noted that there are 'more and more infoboxes in the French Wikipedia which are fully operative with Wikidata'. Despite the advantages, Ange acknowledged that the repository was 'still far from perfect because each and every infobox has to be

converted to the Wikidata way, which is painfully long to do'. To guarantee that the information that features in the infobox is in French, Ange commented that:

> Just after publishing, I make sure that everything is good on Wikidata, because it's really just sad in the end. Because it doesn't take much time to go to Wikidata and just 'Oh, Jane Smith in English, that's Jane Smith in French'. It just takes one second, one click to do it, to add the label in French and... So yeah, and just doing it.

Unlike Ange, Sam (FR) recognised that they did not always translate the information into French. The interviewee added that they 'just put the name and the description [on Wikidata] if [they] can and that's it'. Dominique (FR) noted that one of the highlights of Wikidata was its capacity 'to centralise information into a more easily redistributed form'. Nevertheless, the participant said they had not incorporated it fully into their practice because they disagreed with 'the license the project [had] chosen'.

Maxime (FR), also appreciative of the positive influence that Wikidata had in configuring their practice, was overtly critical of those who raised objections to its deployment:

> Well, honestly, I see two kinds of people that care about Wikidata: the ones that think Wikidata is a great tool and should be used more and should be used more and is very powerful in terms of, well, 'let's use it as it is for the best'; and you have other people who are against Wikidata because, you know, Wikidata is not giving them the opportunity to control everything... So, based on this, we have people who complain all the time because they find that particular problem that is not possible to be fixed because they can't fix it and blah, blah, blah, they won't make the effort of fixing it. So, having stuff in English into French... Yeah, having something written in English in a French infobox in Wikipedia happens. People like me will just fix it; people like them will just complain.

Likewise, Guus (NL) explained that, although 'some people think that the Wikidata interface is too difficult [to edit]', they had found the repository practical and relatively easy to manage. On a par with what Ange said, Jos (NL) observed that the content imported from Wikidata into Wikipedia is not 'at that high level of quality yet'. The participant advocated 'a better editorial process for that kind of content merging' because 'templates don't match up necessarily' and end up putting a strain on translators. Reflecting on their practice, Jos noted that 'translating templates is hard because the fields [in English and Dutch] are slightly different, so

that means a lot of fields get lost in translation'. Leslie (NL) also admitted having problems with templates that 'don't match' and, like their colleague Jos, they called for an improvement to solve 'the technical issues behind it'.

The Swedish Wikipedia participants had mixed views regarding infoboxes imported from Wikidata. Alva noted that one of the main obstacles they had encountered when translating was that sometimes the Swedish Wikipedia had 'more specific templates' for certain subjects that pre-dated the 'more generic' templates imported from Wikidata. When prompted to elaborate on their statement, Alva observed that while the Swedish Wikipedia has specific infoboxes for artists, celebrities and sportspeople, Wikidata classifies them all under the general 'person' template, missing more nuanced achievements within their professions. Robin did not regard this issue as problematic and argued that they preferred to 'take away the infobox' because they believed that translating infoboxes from one language to another 'seldom works well'. In this respect, the interviewee opined that fostering a uniform approach to infoboxes through Wikidata could be optimal for articles on notable people.

Like other participants, Charlie (SV) held that infoboxes imported from Wikidata into Wikipedia still require improvement. Regarding information that has not been translated into the target language, Charlie said that it 'ha[d] never bothered [them]' and criticised fellow editors who resorted to that argument to avoid using Wikidata:

> Yes. A lot of the readers are disturbed when it's not in Swedish. 'It's Swedish Wikipedia; it should be in Swedish'. It has never bothered me; I think the opposite: if I read an article in the Swedish Wikipedia, I usually look for it in the English and German Wikipedia to see if they have said something more. I combine the articles. I have never been disturbed when I see another language. It doesn't bother me; many others are bothered by it and say that's a reason not to use Wikidata templates.

To conclude, the analysis suggests that for some participants, Wikidata has contributed to configuring their translation practices. Although the repository itself figures as infrastructure, it fulfils a similar function to that of a device because it configures the practice of translation in Wikipedia. It is through this configuration that Wikidata acquires an instrumental role akin to those of devices such as bots and CX. Infoboxes imported from the repository have in most cases replaced the ones that were previously managed by individual Wikipedia communities. This

substitution means that, even though infoboxes are now automatically imported into Wikipedia, they may display information that has not yet been translated into the target language. To tackle this issue, some editors may have to devote additional time and effort to editing that data in Wikidata. While this step may come across as a minor change, it is elucidative of how the launch of the repository has configured the translators' practices. Moreover, by translating infobox data from Wikidata into their target languages, Wikipedia editors are also actively shaping the device, hence reinforcing the idea that materials shape practices as much as practices shape materials (Nicolini 2012; Shove, Pantzar, and Watson 2012).

## 6.4 Chapter conclusion

This chapter aimed to ascertain whether and how automation and metadata have contributed to changes in translation practices in Wikipedia over the past lustrum (RQ4). To this end, the analysis focused on data elicited from the last part of the interviews with the 16 participants introduced in Chapter 5. Drawing primarily on Shove's (2017) approach to materials as constituting and configuring elements of practice, the investigation first found that translators in Wikipedia resort to external devices such as online dictionaries, databases and automation to perform their practices and produce texts that meet their expectations.

The data further revealed that opinions differed concerning the mobilisation of bespoke content-creation devices such as bots. The deployment of these tools, subject to strict controls, was generally welcomed by the Spanish Wikipedia participants. However, participants from the Dutch and Swedish Wikipedia communities, which had long relied on bots for content-creation purposes, were divided in their approach to the device. Although some participants were reluctant to use bots on a large scale, others were capable of pondering their benefits, particularly in providing coverage on subjects for which there were not sufficient editors available. Some participants complained that they had to put activities on hold and devote time to correcting bot-generated articles that were basic and of poor quality. Others noted that, once created, bot-generated articles were rarely updated by human editors. Aware of these downsides, the French Wikipedia participants

were among the most sceptical about the deployment of bots, showing an attitude that is aligned with that of their community's shared repertoire.

The vast majority of participants declared that they had incorporated Wikipedia's CX into their practice. According to the interviewees, since its inception in 2014, the device has contributed to configuring their practice by, among other things, reducing the amount of time they used to spend performing manual translations. Despite the advantages, some participants argued that CX was not always efficient in translating metadata such as templates and categories. Unlike other devices such as bots, acceptance of CX was common among participants, although their usage by the Swedish cohort was peripheral. This marginal use of the device among the Swedish Wikipedia participants coincides with the lower number of CX-generated articles in that language community.

Regarding the evolution of translation in Wikipedia, half of the participants believed that new technological advances had configured the practice over the last years. Some were of the view that improvements in automation could eventually lead to a more positive attitude towards machine translation in Wikipedia. Others, however, argued that resistance to automation is still rooted in the encyclopaedia, particularly among senior editors. A small number of participants also contended that Wikipedia has become stricter and that standards of quality have shifted. Alongside changes linked to the mobilisation of automation, a few interviewees envisaged a reconceptualization of translation in the user-driven encyclopaedia with the ongoing development and subsequent launch of Abstract Wikipedia. Such initiative will draw on Wikidata, a repository that has already contributed to configuring translation practices in Wikipedia.

In conclusion, it is apparent from the analysis of the interview data that the deployment of automated devices such as bots and CX has had an impact on the translators' practices across the four Wikipedia communities under investigation. On the one hand, the Dutch and Swedish Wikipedia versions have benefited from, and have been affected by, the mobilisation of bots. On the other, the Spanish and French communities have seen their number of CX-generated articles increase in the past years, even more so since McDonough Dolmaya's (2017) study. Along with these devices, the growing influence of repositories such as Wikidata has

enabled the importation of the same metadata into individual language communities of Wikipedia. However, parts of the metadata displayed on infoboxes are not always translated into the target languages, hence forcing some editors to devote additional time to updating the information once the articles have been posted on Wikipedia. In light of the impact of Wikidata in configuring the translators' practices, it can be argued that the repository fulfils a role akin to internal devices such as bots and CX.

# Conclusion

This last chapter of the thesis aims to bring together the primary findings of the investigation and discuss them within the broader context of Wikipedia's multilingualism. To begin with, 7.1 will address the four research questions in separate subsections (7.1.1-7.1.4). Section 7.2 will then move on to reflect on the main contributions of the study. Finally, 7.3 will suggest avenues for future research.

## 7.1 Evolution of translation practices over the last lustrum

This doctoral thesis has investigated how regulation, negotiation and automation have contributed to the evolution of translation practices in four language communities of Wikipedia over the past five years. In Chapter 1, it was argued that the limited research on translation in the user-driven encyclopaedia (McDonough Dolmaya 2012; 2015; 2017; Jones 2018b; Shuttleworth 2018) has, for the most part, overlooked the role of local translation guidelines and automated devices in informing and configuring the translators' practices. Moreover, despite the growing interest in Wikipedia's multilingualism throughout the last decade (Callahan and Herring 2011; Ensslin 2011; Lewoniewski, Węcel, and Abramowicz 2018; Lewoniewski 2019; Park et al. 2020), little attention has been paid to the significance of cross-lingual content-creation devices in aiding and enabling the process of peer co-production. Barring some notable exceptions (Laxström, Giner, and Thottingal 2015; McDonough Dolmaya 2017), the impact of tools such as CX remains understudied. Likewise, research on bot deployment in Wikipedia has been sparse (Geiger 2011; Halfaker and Riedl 2012; de Laat 2015) and has failed to ascertain the influence of the device in the creation of multilingual encyclopaedic material.

Against this backdrop, Chapter 2 introduced practice-theoretical concepts to better understand the standards and devices that underpin the performance of translation in Wikipedia. Drawing primarily on Wenger's (1998) concept of 'communities of practice' and Warde's (2005; 2016) 'standards of performance', chapters 4 and 5 of this research project sought to examine the regulation and negotiation of translation guidelines and bot-creation policies in the Spanish, French, Dutch and

Swedish language communities of Wikipedia. As discussed in 2.3, each of the four language versions of the user-driven encyclopaedia was approached as a separate community of practice. This classification was made based on the documented existence of distinct values, concepts, and standards collectively known as the 'shared repertoire', which regulate translation and restrict access to specific bespoke automated devices such as bots and CX.

As explained in 2.1.1 and 2.2.1, arguably some Wikipedia policies such as Verifiability (WP: VER), Notability (WP: NOTE), Neutral Point of View (WP: NPOV), and Copyright (WP: COPY) can be regarded as 'universal' on the grounds that they are shared by most language communities of the encyclopaedia. Although there has been ample research on the violation of core policies such as NPOV (Callahan and Herring 2011; Góngora-Goloubintseff 2020; Oeberst et al. 2020), compliance with these policies is expected if one wants to become a full-fledged editor in Wikipedia. An elucidative example can be found in the introductory section of NPOV: 'This policy is *non-negotiable*[43], and the principles upon which it is based cannot be superseded by other policies or guidelines, nor by editor consensus' ('Wikipedia:Neutral Point of View' 2021).

While editing is explicitly and universally regulated in Wikipedia, there is less cross-lingual consensus concerning standards of translation performance. This division is further attested by the fact that none of the documents analysed in 4.1 and 4.2 transcend the confines of their local Wikipedia communities. Moreover, the lower 'guideline' status of most of these standards indicates that they are not strictly implemented. In the case of the Swedish Wikipedia, where the 'translation tips' do not even enjoy such status, compliance with their precepts seems to be left at the discretion of individual translators.

Regardless of their locality and status, the existence of these documented standards is evidence of a concerted effort to regulate translation practices in certain communities of Wikipedia. To better understand how these practices have emerged and evolved, this thesis first set out to investigate their regulation and negotiation (Chapter 4) and then moved on to examine their incorporation by 16 experienced editors (Chapter 5). Finally, to complete the study of how translation practices have

[43] My emphasis.

changed, the analysis conducted in Chapter 6 tackled the impact of content-creation devices and metadata, focusing on the past five years. The following subsections provide a summary of the main findings of the research.

### 7.1.1 Regulation of translation practices: the standards

The first part of Chapter 4 was articulated around RQ1: 'How and to what extent have the four Wikipedia language communities regulated translation practices?' As explained in 3.2.1, to answer this question, a thematic analysis was conducted using data gathered from documented standards of translation and bot policies. Drawing on Wenger's (1998) 'communities of practice' and Warde's (2005) 'standards of performance', the investigation found that, despite the existence of perceptible differences in the advice they give to translators, all four language versions of Wikipedia share a series of overarching principles on translation. Specifically, 4.1 unearthed the existing 'overlap' concerning verifiability of content and acknowledgement of the sources, automatic translations, prescriptive expectations around 'language correctness', and guidance on using CX.

Although this overlap can be used to frame multilingual Wikipedia as a constellation of communities of practice, such convergence is also illustrative of the lack of clear-cut boundaries between editing and translation in the user-driven encyclopaedia. As discussed in 1.1.1, previous research by Jones (2017; 2018b) already argued that translation in Wikipedia is best described as another form of editing or original writing. Likewise, Shuttleworth (2018, 234) noted that, due to Wikipedia's volatility, where most articles undergo changes over time, it may at times be challenging to draw the line between translation and editing. By shifting the attention to the study of translation standards, this thesis has yielded findings consistent with those claims. In particular, the first part of the analysis showed that non-negotiable editing policies such as Verifiability (WP: VER) feature prominently in the documented standards under investigation.

Regarding points of dissonance, the data analysis indicates that the four communities have set a series of clearly distinct recommendations that translators should attempt to follow. For instance, while the Spanish and French Wikipedia guidelines recommend selecting an article for translation provided it meet certain

high-quality standards, the Dutch Wikipedia document places more value on the editor being familiar with the topic covered in the ST. Moreover, in 4.2.3, it became apparent that the Dutch Wikipedia guidelines are markedly different in discouraging translators from importing references they have not consulted into the TT. As observed in 4.2.4, this distrust of the source is also found in the Swedish Wikipedia essay, where a cooking analogy is employed to remind editors that caution is necessary when approaching English Wikipedia articles.

Another significant difference that surfaced during the analysis is the gross disparity between the four communities concerning policies on bot-generated content. For instance, a close reading of the documents revealed that neither the Dutch nor the Swedish versions established any provisions. This paucity of regulation was later corroborated by a few editors of those communities, who were contacted via email correspondence. Conversely, the Spanish Wikipedia community approved a policy in 2008 to restrict the deployment of bots for content-creation purposes. As analysed in 4.2.3 and 4.3.1, editors of that community were divided regarding the approval of these stringent measures. Equally restrictive, the French Wikipedia occupies a place in-between. Like the Dutch and Swedish communities, the French Wikipedia lacks documented standards regulating bot-generated content. However, the data show that the community is not generally welcoming of bot creations (‘Wikipedia Statistics’ 2019).

As explained in 2.3, despite having a lower status than policies, guidelines in Wikipedia remain ‘sets of practices supported by consensus’ (‘Wikipedia:Policies and Guidelines’ 2021). From a praxeological lens, consensus or mutual engagement is necessary for the emergence and long-term sustainment of practice (Wenger 1998; Shove, Pantzar, and Watson 2012; Warde 2016). At the same time, this thesis has contended that the existence of cross-wiki differences around how translation should be performed further lends support to the idea that each language community of Wikipedia has a unique shared repertoire. Building on the analysis conducted in chapters 4 to 6, the following three subsections will discuss in more detail whether this shared repertoire is preserved or challenged by editors and experienced translators of the four communities. In doing so, the objective is to ascertain the impact of either action on the evolution of translation in Wikipedia.

### 7.1.2 Negotiation of the standards

Following the examination of the documented standards, attention was then turned to their negotiation in the ancillary discussion forums known as 'talk pages'. Thus, the second part of Chapter 4 sought to answer RQ2: 'How are translation standards negotiated in each of the four language communities?' As observed by Wenger, McDermott, and Snyder (2002), online forums play a significant role in communities of practice whose members work remotely, mainly because such spaces contribute to developing a sense of togetherness. In the context of a linguistically and culturally diverse user-driven encyclopaedia such as Wikipedia, talk pages are essential communication channels where editors can openly negotiate aspects of the shared repertoire by voicing their concerns, asking questions, putting forward changes to documents and articles, and starting discussion threads.

As seen in 1.1.1, scholars such as Jones (2017; 2018b; 2018a) have probed into talk pages to better understand how translation narratives unfold in Wikipedia articles. His research found that translators are not mere altruistic bridges, as previous studies suggested (McDonough Dolmaya 2012) but occasionally compete against one another. Against this backdrop, this thesis further investigated the role of talk pages in the negotiation of documented standards of translation. The thematic analysis of the data showed that negotiations around translation standards in the four communities tended to be lengthy, scattered over time, and centred on specific sections of the documents. Moreover, in line with Jones' (2018a, 2018b) findings, the investigation revealed instances of conflict and hostility among the parties involved, particularly in the Spanish Wikipedia debate on bot-generated content (4.3.1), the French Wikipedia discussion on automatic translations (4.3.2), and the Dutch Wikipedia thread on references (4.3.3).

Considering the size and diversified range of users spread across Wikipedia communities, disagreements among volunteers are not only commonplace, but they may ultimately contribute to shaping practices themselves. As Wenger (1998) posits, engaged members of a community of practice may disagree with one another in the pursuit of their joint enterprise. This 'rebellion' often transpires when some aspects of the shared repertoire are challenged. Similarly, Warde (2005, 141)

observes that ‘conventions will to some extent be contested’, with senior practitioners often unwilling to embrace new ways of doing. In this study, the generational divide between newcomers and seniors was not addressed, but the examination uncovered instances of friction among experienced users.

One elucidative example is found in the Spanish and French Wikipedia discussions on the deployment of content-creation devices. In the former, the 2008 debate on whether bot-generated content should be allowed was met with the outward hostility of a small group of engaged editors who viewed those tools as a threat to the community. Although the policy was approved in the end, the number of bot-generated articles in the Spanish Wikipedia remains scant. In the French Wikipedia, where the community decided to draft a policy to regulate automatic translations before the launch and subsequent incorporation of CX, criticism arose when some editors perceived the move as an attempt to ‘demonise’ automation. As Chapter 6 examined and as 7.1.4 will readdress, the division between those who endorse automation and those who oppose it has not waned over the years. Some Dutch and Swedish Wikipedia participants, known for lacking regulation and being more permissive, were also against mobilising bots.

Overall, the analysis conducted in Chapter 4 suggests that editors tend to have particular areas of concern. Most discrepancies revolved around editing policies featuring prominently in the standards like WP:VER, but also license agreements and the mobilisation of devices such as bots and automatic translators. The data gathered from the talk pages further indicate that negotiations do not always result in changes to the standards and that little activity has taken place over the past five years. The preliminary findings implied that, on observing the limited success of past discussions, some editors might have decided to avoid entering into taxing debates. Nevertheless, it was also hypothesised that low levels of activity around the standards could either be an indication of widespread compliance with their provisions (passive conformity) or evidence of the small weight that guidelines and essays have in informing translation practices in Wikipedia. To ascertain whether any of these hypotheses was correct, the third stage of the analysis targeted data collected from semi-structured interviews with 16 experienced Wikipedia translators.

### 7.1.3 Incorporation of the standards

Building on the findings from the two previous analytical stages, this part of the investigation addressed RQ3: 'To what extent have experienced Wikipedia translators incorporated the standards set by their language communities into their practices?' To this end, and as was explained in 3.4.1, a total number of 16 Wikipedia translators – four per language community – were selected following a series of purposive sampling criteria. Some of the participants' answers echo the findings from earlier studies that investigated in greater length the motivations driving volunteer translators (O'Brien and Schäler 2010; McDonough Dolmaya 2012; Olohan 2014; Cámara de la Fuente 2015). Moreover, the data analysed in 5.3 show that the participants had to overcome a series of common challenges along their inbound trajectory from newcomers to full-fledged practitioners. Admittedly, some of these obstacles, particularly regarding wikicode, have also been reported in previous studies (Lih 2009; O'Hagan 2016) and can be broadly divided into categories: a) issues stemming from the ST (mostly lexicon-related), and b) difficulties concerning aspects of wikisyntax such as complex coding, learning how to run the Visual Editor, and understanding Wikipedia categories and infoboxes.

When it comes to the incorporation of community-specific translation standards, most participants were overtly sceptical about their relevance not only to their own practice but also to that of their colleagues. While further research is necessary to ascertain whether this latter claim is valid, the interviewees' answers indicate that there may be some robust reasons behind the inconsequentiality of the documented standards. For instance, recurrent themes that emerged during the analysis of the data were a) the lack of currency and optionality of translation guidelines, b) the fact that they are not easy to find, and c) the perceived gap between the advice they sometimes give (what practitioners ought to achieve) and what translators actually do. Expanding on this last argument, a few participants – particularly from the Dutch Wikipedia – were critical of the importance that some editors seem to attach to guidelines.

Despite the widespread tendency among participants to pay little if any attention to translation standards, their responses indicate that a vast majority has managed to

incorporate core policies and values into their practice. Such policies and values, while cited in the translation guidelines, tend to be overarching, non-negotiable, and common to other forms of editing. To put it more concretely, the data suggest that participants of the four communities prioritise obligatory cross-wiki editing policies over optional local translation standards. Perhaps the most elucidative example is Verifiability (WP: VER), which was mentioned by all interviewees. Other policies, in declining order of preference, were Notability (WP: NOTE) and Neutral Point of View (WP: NPOV). To some extent, all three editing policies feature in the documented standards of translation. Their inclusion demonstrates that local translation guidelines draw on editing policies and that the same rules apply to all Wikipedia editors. These expectations are not unrealistic considering that the quintessential Wikipedia article should comply with the project policies, regardless of how it was created.

In addition to providing evidence that situates translation in Wikipedia as a subtype of editing, the analysis also revealed that a considerable number of participants had incorporated a series of principles outlined in the documented standards. The fact that this incorporation transpired without consultation of translation standards shows that certain aspects of the practice are implicitly performed and that little regard for and marginal exposure to the documented standards does not necessarily result in non-compliance. For example, the recommendation that editors should select articles for translation only if they are familiar with the topic is explicitly stated in the Dutch Wikipedia guidelines (4.2.1), but it is not mentioned elsewhere. Yet, four participants of other Wikipedia communities declared using familiarity as a criterion for choosing their articles.

As was observed in 5.5 and as will be discussed in more detail in 7.2, the participants' sustained exposure to the practice of translation in Wikipedia may have contributed to the assimilation of explicit skills. Simply put, through the repeated performance of translation in the user-driven encyclopaedia, a few participants are likely to have learned and acquired specific knowledge and expectations about their practice, which have now become implicit to them. Another possible interpretation is that some participants may be acting as brokers (Wenger 1998). This explanation does not cancel the previous one; instead, it suggests that some participants could have gained their knowledge from outside

their primary Wikipedia language communities. Although further research is required to ascertain whether that is the case, the data analysis conducted in chapters 5 and 6 indicates that a few participants were able to make judgments about what transpired in other language communities, either as outsiders or as members. The next subsection will expand on this hypothesis when addressing cross-wiki differences regarding the mobilisation of automated devices.

### 7.1.4 Changes stemming from automation

Drawing upon the same interview data that backed up the findings of Chapter 5, the previous chapter aimed to answer RQ4: 'How and to what extent have automation and metadata contributed to changes in translation practices in Wikipedia over the last five years?' To gain a better understanding of how materials configure practices, the analysis undertaken in Chapter 6 brought to the foreground Shove's (2017) concept of 'devices'. First discussed in 2.4, this tripartite framework states that devices exist in a relationship of continued interdependence with resources and infrastructure (Shove 2017, 160). Unlike the last two, however, the mobilisation of devices plays a more fundamental role in configuring practice. For example, as was explained in 2.4, practices such as tennis and driving have evolved throughout history thanks in part to substantial material enhancements (Shove, Pantzar, and Watson 2012).

Although there has been some research on the impact of materials on translation (Littau 2016; Olohan 2017; 2021), previous studies on Wikipedia have overlooked the importance of content-creation devices such as CX and bots for the performance and evolution of the practice. As shown in 1.1.1, apart from a study by McDonough Dolmaya (2017), where CX is briefly mentioned, most research has centred the attention elsewhere. Thus, this thesis has sought to gain better insight into how individual Wikipedia language communities regulate and negotiate the deployment of automation (Chapter 4) and the extent to which automated devices and metadata have contributed to the evolution of translation in the online encyclopaedia (Chapter 6). Focusing the analysis on the last five years, roughly coinciding with the launch of CX, the investigation found that most participants had incorporated CX into their practice. As became apparent in 6.2.2, this device configured how the participants performed translation, primarily by making the process faster and smoother, but

also by simplifying other tedious tasks such as acknowledging the source of the translated article.

The participants' answers analysed in the previous chapter suggest that the mobilisation of CX has had a perceptible impact on their translation practices. Besides optimising their time, the answers given by some of the interviewees, particularly from the Dutch Wikipedia, show that regular software updates may have lured the attention of both CX-sceptical and inexperienced Wikipedia translators. These upgrades, however, have yet to resolve issues of incompatibility affecting the translation of templates and infoboxes, as well as problems commonly found in MT, such as inaccurate use of language.

Another consequence of the mobilisation of automated devices in Wikipedia is what some participants perceived to be a lack of thematic diversity in the articles. As observed in 1.1.1, a previous study (McDonough Dolmaya 2017) has hypothesised that an overreliance on the English Wikipedia as the primary source of the translations coupled with the lack of an official translation policy in the encyclopaedia could result in a large number of translated articles containing one-sided information. In Chapter 6, it became apparent that a few participants had similar concerns amid the rapid proliferation of CX-generated articles. This study also found that the deployment of other content-creation devices such as bots led to even more mixed reactions among members of the four Wikipedia communities.

Participants of the Spanish and French Wikipedia, communities known for having restricted the creation of articles with bots, expressed diametrically opposed views. The former showed a more positive attitude towards bots, especially when compared to the views held by some of their colleagues in the 2009 debate (4.3.1). The latter, having witnessed the disruption caused by Lsjbot in the Swedish Wikipedia, were more reluctant to mobilise the device in their community. The French participants' position as outsiders or members of more than one Wikipedia (brokers) seems to have contributed to reaffirming their stance on bot-generated content, which is in tune with the values that make up their community's shared repertoire. Viewpoints were more varied among participants of the Dutch and Swedish Wikipedia communities, most of whom had come across bot-generated articles. Moreover, although none of them reported owning a bot for such purposes,

a few indicated that they had devoted additional time and efforts to revising and amending articles created by the device.

To conclude, the analysis showed that even though internal devices such as CX have played a central role in configuring most participants' practices in the last lustrum, the mobilisation of external aids such as online dictionaries, translation databases, and MT has helped some Wikipedia translators achieve their goals (6.1). Besides, the inception of the Wikidata platform in 2012 marked a turning point for Wikipedia translation. The repository, which is the result of a process of centralisation of knowledge, has taken over some tasks that Wikipedia translators used to perform manually (6.3.1). However, as was the case with CX and bots, the incorporation of Wikidata also came at a cost, with some participants complaining that infoboxes are not always rendered correctly into the target languages. The effort invested in solving these issues and the fact that to do so Wikipedia editors are expected to acquire new skills lend support to the idea that the repository, an example of infrastructure, can function as a device. As discussed in 2.4 and Chapter 6, Shove, Pantzar, and Watson (2012), Hui (2017), and Shove (2017) already observed that materials could take on different roles depending on the practice they configure. The information currently available and the participants' comments suggest that the future launch of projects such as Abstract Wikipedia is likely to bring about even more radical changes to the practice.

## 7.2 Contribution of the study

This thesis has contributed to enhancing the understanding of the standards and materials that underpin translation practices in collaborative, multilingual projects such as Wikipedia. By targeting four language communities based on perceptible cross-wiki differences in the deployment of automated devices, the first aim of this study was to ascertain the relevance of local documented standards (policies, guidelines and essays) in informing the practice of translation in the online encyclopaedia. Closely related to this aim, the second objective was to establish whether and if so how automated devices and metadata have played a significant role in configuring the practice over the past five years.

Similar to previous research on translators as a community of practice (Neather 2012; Yu 2019), this study showed that editor-translators in Wikipedia mutually engage to propose amendments to documented standards. Such standards, in the form of policies, guidelines and essays, are part of the Wikipedia community's shared repertoire. Although compliance with the guidelines and essay on translation is ultimately a matter of personal choice, such documents are illustrative of how individual communities have attempted to establish boundaries between translation and editing in Wikipedia. Another plausible explanation deriving from the analysis is that translation guidelines or 'help pages' function as recruiting agents. Drafted and approved by senior practitioners, guidelines seek to socialise peripheral members and outsiders into the basic standards underpinning translation in a particular Wikipedia community. To this end, they give advice on what to do to become a full-fledged practitioner and avoid instances of conflict with other members. Thus, observance of these rules, while discretionary, is at a minimum required to guarantee a successful inbound trajectory to membership in the online encyclopaedia.

Perhaps because of their optionality, the selected documents have not undergone substantial changes, neither have they been widely contested. Instances of rebellion among editors were limited to particular aspects of the standards such as referencing, style and the use of automation. Despite this, a few discussion threads presented clear instances of vehement opposition from a reduced yet boisterous group of editors. The contentious nature of some of these discussions, consistent with findings from previous studies (Jones 2017; 2018b; 2018a), could have deterred other editors from engaging in the conversation. This tendency was observed in all four communities, suggesting widespread passive conformity, a general lack of interest in updating the documents or both. In the case of the Spanish Wikipedia policy on bot-generated content, the negotiation preceding its approval was conspicuously more active, with two differentiated sides holding almost irreconcilable views. Nevertheless, the rebellion was short-lived, and little has transpired since the debate ended in early 2008.

The Wikipedia translators that took part in the study soon corroborated that the prolonged period of inactivity in the discussion threads was likely to be the result of several factors. The most notable of these was the marginal role that local

translation standards played in informing their practice. Although the cohort was small – as only four participants per language community were selected – the responses provide a good indicator of the low priority that documented standards had for those interviewed. Once more, there were no meaningful differences between participants of individual Wikipedia communities in how they approached the translation standards. Commonly raised issues around these pages were the vacuous advice they sometimes give, especially regarding aspects such as referencing and their inability to meet current demands. It is quite probable that the protracted idleness reported in the talk pages may have a bearing on the lack of currency of the standards. However, further research is necessary to unearth the underlying reasons as to why members of the four Wikipedia communities have not invested more time in updating the documents.

Despite the fact that documented standards have not undergone considerable changes in recent years, the participants' answers demonstrate that translation in Wikipedia has evolved over the last lustrum. Partly due to the inception of CX and the tightening of editing policies, the translation process seems to have become quicker, stricter and more reliant on automation. While most participants agreed that bespoke devices such as CX were far from perfect, the vast majority recognised the positive shaping effect these materials had on their practice. A second insight, arising from the participants' answers on the mobilisation of bots, is that these content-creation devices, first launched to fill knowledge gaps, were generally perceived as more potentially disruptive. To elaborate, a few participants of the Dutch and Swedish Wikipedias indicated that their communities are still dealing with the consequences of the mishandling of bots. As became apparent in the analysis, participants of the French Wikipedia were not ignorant about the long-lasting effects of poorly written bot-generated content in the communities that had been more permissive of their deployment. In parallel, the Spanish Wikipedia interviewees, slightly more supportive of bots, were in favour of mobilising the device subject to scrutiny, as established in the 2008 policy.

Finally, the fact that most participants were able to comment on overarching editing policies but showed little awareness of local translation standards indicates that the major enforceability of the former, as opposed to the optionality of the latter, may be one of the reasons behind their prioritisation. Nevertheless, the investigation

found evidence that the vast majority of participants had implicitly incorporated some of the advice given in the standards. From a praxeological perspective, there are at least four possible explanations for this incorporation, none of which are mutually exclusive. First, participants could have consulted the documents at some point during the course of their inbound trajectory to membership. Second, participants may have gained implicit knowledge through performing their practice without resorting to standards. Third, participants are likely to have become competent translators by following the same standards that underpin editing practices in Wikipedia. Fourth and last, by virtue of being members of other Wikimedia projects, including other Wikipedia communities, participants could be acting as brokers, importing skills acquired elsewhere into their primary language community.

## 7.3 Concluding remarks and suggestions for future research

This study examined translation in Wikipedia from a practice-theoretical perspective. The findings emerging from this thesis show that, despite there being distinct local documented standards on translation, the core set of values, expectations and materials that underpin and configure the practice are similar across the Wikipedia communities under investigation. Most participants coincided on their adherence to overarching editing policies as well as on the little importance they attached to translation standards. Thus, this research project has found evidence that lends support to previous assertions that translation and editing in the user-driven encyclopaedia form a continuum (Shuttleworth 2017; Jones 2018b).

Moreover, while negotiation over the translation standards was limited and scattered in time, the heated debates around particular aspects of the documents indicate that dispute and confrontation may be part of the change process. Although other studies already uncovered instances of disagreement and tension among Wikipedia translators (Jones 2017), the practice-theoretical lens adopted in this thesis allows for an interpretation of conflict or rebellion as a driving force of change, stemming from the practitioners' strong sense of commitment to improving their community's shared repertoire.

Alongside regulation and negotiation of the standards, this study has sought to ascertain the extent to which materials both configure and reconfigure translation practices in Wikipedia. In particular, the research undertaken in this thesis centred the attention on the pivotal role of materiality in configuring the practices of volunteer translators, an area that remains understudied (Littau 2016; Olohan 2014, 2017, 2021). Consequently, this thesis drew upon conceptual elements from Shove (2017) to assess the impact of multilingual content-creation devices on the practices of translators in Wikipedia. One distinctive aspect of this study is that the analysis of automation was not merely limited to translation tools but also included bots. These tools were considered along with CX because of their long track record of being deployed to facilitate and accelerate the dissemination of knowledge across various language communities of the user-driven encyclopaedia.

Despite their imperfections, automated devices such as CX and bots have been mobilised in Wikipedia mostly because of their proven ability to optimise the translator-editors' performance. As this study has demonstrated, bots have been deployed to fill knowledge gaps, especially in language communities where there is a lack of volunteers available or willing to take on the task. Other devices more closely associated with translation, such as CX, have been launched with a similar aim in mind: to disseminate knowledge across language communities as quickly and effectively as possible. By allowing editors to work in a user-friendly format, where they are no longer expected to deal with intricate wikicode, CX has significantly contributed to reducing their workload and improving their productivity.

In essence, the desire to be effortlessly more productive in achieving goals is something that has driven changes in practice throughout history. In the case of Wikipedia, as this study has shown, the mobilisation of more user-friendly and appealing devices stems from a communal need to recruit new practitioners. Through this recruitment process, each language community can sustain a healthy number of members capable of working together towards achieving the joint enterprise. The fulfilment of this enterprise already became apparent in Jimmy Wales' 'prime objective', but it has recently become more evident with the ongoing development and projected launch of Abstract Wikipedia. If, as has been suggested, code will do the translation, it would be interesting to see how human editors would

fit in that scenario. Thus, one fruitful avenue of research could lie in the examination of the impact that such an initiative could have not only in configuring but also in potentially reconceptualising translation in collaborative projects such as the online encyclopaedia. Based on the findings of the investigation, however, it seems reasonable to conclude that human translators will continue to engage and remain essential agents in the dissemination of knowledge for years to come. The discontinuation of bots in some language communities and the rapid increase in the number of translators deploying CX in others show that the practice is gathering momentum in a twenty-year-old encyclopaedia whose success has long resided in the fact that 'anyone can edit'.

# Bibliography

## 8.1 Primary sources

Below are listed all the Wikipedia pages (articles and documents) selected for the study, as well as other online sources that were consulted during the investigation.


‘Abstract Wikipedia’. 2021. Meta-Wiki. 11 August 2021. https://meta.wikimedia.org/wiki/Abstract_Wikipedia.

‘Aide:Outil de traduction’. 2020. Wikipedia FR. 14 April 2020. https://fr.wikipedia.org/wiki/Aide:Outil_de_traduction.

‘Aide:Traduction’. 2020. Wikipedia FR. 24 May 2020. https://fr.wikipedia.org/wiki/Aide:Traduction.

‘Användardiskussion:Lsj’. 2020. Wikipedia SV. 1 May 2020. https://sv.wikipedia.org/wiki/Anv%C3%A4ndardiskussion:Lsj.

‘Användare:Lsj’. 2017. Wikipedia SV. 20 July 2017. https://sv.wikipedia.org/wiki/Anv%C3%A4ndare:Lsj.

‘Ayuda:Cómo traducir un artículo’. 2020. Wikipedia ES. 24 June 2020. https://es.wikipedia.org/wiki/Ayuda:C%C3%B3mo_traducir_un_art%C3%ADculo.

‘Content Translation’. 2020. MediaWiki. 31 May 2020. https://www.mediawiki.org/wiki/Content_translation.

‘Content Translation Statistics’. 2021. Wikipedia. 11 August 2021. https://en.wikipedia.org/wiki/Special:ContentTranslationStats.

‘CZ:Policies’. 2020. Citizendium. 2020. https://en.citizendium.org/wiki/CZ:Policies.

‘Discussion aide:Outil de traduction’. 2020. Wikipedia FR. 17 March 2020. https://fr.wikipedia.org/w/index.php?title=Discussion_aide%3AOutil_de_traduction&type=revision&diff=168508978&oldid=168508265.

‘Discussion aide:Traduction’. 2020. Discussion thread. *Wikipedia FR*. https://fr.wikipedia.org/w/index.php?title=Discussion_aide%3ATraduction&type=revision&diff=170976820&oldid=170976748.

‘Discussion Wikipédia:Traduction automatique’. 2020. *Wikipedia FR*. https://fr.wikipedia.org/w/index.php?title=Discussion_Wikip%C3%A9dia%3ATraduction_automatique&type=revision&diff=167347507&oldid=167336541.

‘Help:Creating a Bot’. 2021. Wikipedia. 9 July 2021. https://en.wikipedia.org/wiki/Help:Creating_a_bot#Programming_languages_and_libraries.

‘Help:Gebruik van bots’. 2020. Wikipedia NL. 28 April 2020. https://nl.wikipedia.org/wiki/Help:Gebruik_van_bots.

‘Help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia’. 2019. Wikipedia NL. 21 August 2019. https://nl.wikipedia.org/wiki/Help:Tips_voor_het_vertalen_van_een_artikel_vanaf_een_andere_Wikipedia.

‘Help:Translation’. 2021. Wikipedia. 3 August 2021. https://en.wikipedia.org/wiki/Help:Translation.

‘Lev Levin’. 2021. Wikipedia ES. 22 August 2021. https://es.wikipedia.org/w/index.php?title=Lev_Levin&oldid=137834332.

‘Lsjbot’. 2021. Wikipedia. 30 July 2021. https://en.wikipedia.org/wiki/Lsjbot.

‘Overleg help:Tips voor het vertalen van een artikel vanaf een andere Wikipedia’. 2015. *Wikipedia NL*.

https://nl.wikipedia.org/wiki/Overleg_help:Tips_voor_het_vertalen_van_een_artikel_vanaf_een_andere_Wikipedia.

‘Portal:Huvudsida’. 2021. Wikipedia SV. 14 August 2021. https://sv.wikipedia.org/wiki/Portal:Huvudsida.

‘Q230899’. 2021. Wikidata. 27 June 2021. https://www.wikidata.org/w/index.php?title=Q230899&action=history.

‘Theresa May’. 2021. Wikipedia. 10 August 2021. https://en.wikipedia.org/wiki/Theresa_May.

‘Viquipèdia:Categorització’. 2021. Wikipedia CA. 2 May 2021. https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Categoritzaci%C3%B3#Biografies.

‘Votaciones:Botopedia’. 2008. Wikipedia ES. 20 January 2008. https://es.wikipedia.org/wiki/Wikipedia:Votaciones/2008/Creaci%C3%B3n_de_art%C3%ADculos_con_bots.

Wales, Jimmy. 2021. ‘Statement of Principles’. Wikipedia. 20 March 2021. https://en.wikipedia.org/wiki/User:Jimbo_Wales/Statement_of_principles.

‘Wikidata Main Page’. 2021. Wikidata. 12 August 2021. https://www.wikidata.org/wiki/Wikidata:Main_Page.

‘Wikipedia’. 2021. Wikipedia. 10 August 2021. https://en.wikipedia.org/wiki/Wikipedia.

‘Wikipedia discusión:Votaciones/2008/Creación de artículos con bots’. 2008. *Wikipedia ES*. https://es.wikipedia.org/wiki/Wikipedia_discusi%C3%B3n:Votaciones/2008/Creaci%C3%B3n_de_art%C3%ADculos_con_bots.

‘Wikipedia Statistics’. 2019. Wikimedia. 31 July 2019. https://stats.wikimedia.org/EN/BotActivityMatrixCreates.htm.

‘Wikipedia Talk:Translate Us’. 2020. *WikiProject:Translation Studies*. https://en.wikipedia.org/wiki/Wikipedia_talk:Translate_us.

‘Wikipédia:Accueil_principal’. 2021. Wikipedia FR. 14 August 2021. https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal.

‘Wikipedia:Bots’. 2021. Wikipedia. 11 April 2021. https://en.wikipedia.org/wiki/Wikipedia:Bots.

‘Wikipedia:Content Translation Tool’. 2021. Wikipedia. 25 July 2021. https://en.wikipedia.org/w/index.php?title=Wikipedia:Content_translation_tool&action=history.

‘Wikipedia:ContentTranslation’. 2018. Wikipedia NL. 25 January 2018. https://nl.wikipedia.org/wiki/Wikipedia:ContentTranslation.

‘Wikipedia:Creación de artículos con bot’. 2020. Wikipedia ES. 9 February 2020. https://es.wikipedia.org/wiki/Wikipedia:Creaci%C3%B3n_de_art%C3%ADculos_con_bot.

‘Wikipedia:Creación de artículos con bot/Solicitudes’. 2019. Wikipedia ES. 7 September 2019. https://es.wikipedia.org/wiki/Wikipedia:Creaci%C3%B3n_de_art%C3%ADculos_con_bot/Solicitudes.

‘Wikipédia:Créer un bot’. 2020. Wikipedia FR. 20 April 2020. https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Cr%C3%A9er_un_bot.

‘Wikipediadiskussion:Översättningsrekommendationer’. 2018. *Wikipedia SV*. https://sv.wikipedia.org/wiki/Wikipediadiskussion:Rekommendationer_vid_%C3%B6vers%C3%A4ttning_fr%C3%A5n_engelska.

‘Wikipedia:Five Pillars’. 2021. Wikipedia. 11 June 2021. https://en.wikipedia.org/wiki/Wikipedia:Five_pillars.

‘Wikipedia:Harassment’. 2021. Wikipedia. 4 July 2021. https://en.wikipedia.org/wiki/Wikipedia:Harassment.

‘Wikipedia:Mentorship’. 2020. Wikipedia. 4 October 2020. https://en.wikipedia.org/wiki/Wikipedia:Mentorship.

‘Wikipedia:Neutral Point of View’. 2021. Wikipedia. 6 August 2021. https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view.

‘Wikipedia:No Original Research’. 2021. Wikipedia. 18 July 2021. https://en.wikipedia.org/wiki/Wikipedia:No_original_research.

‘Wikipedia:Notability’. 2021. Wikipedia. 2 July 2021. https://en.wikipedia.org/wiki/Wikipedia:Notability.

‘Wikipedia:Opinielokaal/Botimport’. 2011. Wikipedia NL. 18 December 2011. https://nl.wikipedia.org/wiki/Wikipedia:Opinielokaal/Botimport.

‘Wikipedia:Översättningsrekommendationer’. 2020. Wikipedia SV. 8 April 2020. https://sv.wikipedia.org/wiki/Wikipedia:Rekommendationer_vid_%C3%B6vers%C3%A4ttning_fr%C3%A5n_engelska.

‘Wikipedia:Policies and Guidelines’. 2021. Wikipedia. 3 August 2021. https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines.

‘Wikipedia:Portada’. 2021. Wikipedia ES. 14 August 2021. https://es.wikipedia.org/wiki/Wikipedia:Portada.

‘Wikipedia:Prime Objective’. 2021. Wikipedia. 22 January 2021. https://en.wikipedia.org/wiki/Wikipedia:Prime_objective.

‘Wikipedia:Rekommendationer’. 2021. Wikipedia SV. 16 July 2021. https://sv.wikipedia.org/w/index.php?title=Wikipedia%3ARekommendationer_vid_%C3%B6vers%C3%A4ttning_fr%C3%A5n_engelska&type=revision&diff=49452044&oldid=49394798.

‘Wikipedia:Robotar’. 2019. Wikipedia SV. 10 February 2019. https://sv.wikipedia.org/wiki/Wikipedia:Robotar.

‘Wikipédia:Traduction automatique’. 2018. Wikipedia FR. 21 October 2018. https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Traduction_automatique.

‘Wikipedia:Upphovsrätt’. 2020. Wikipedia SV. 9 April 2020. https://sv.wikipedia.org/wiki/Wikipedia:Upphovsr%C3%A4tt.

‘Wikipedia:Verifiability’. 2021. Wikipedia. 30 July 2021. https://en.wikipedia.org/wiki/Wikipedia:Verifiability.

‘Wikipedia:Votaciones finalizadas’. 2008. Wikipedia ES. 2008. https://es.wikipedia.org/wiki/Categor%C3%ADa:Wikipedia:Votaciones_finalizadas_2008.

‘Wikipedia:Votaciones/2008/Sustitución de la política de botopedia’. 2018. Wikipedia ES. 31 May 2018. https://es.wikipedia.org/wiki/Wikipedia:Votaciones/2008/Sustituci%C3%B3n_de_la_pol%C3%ADtica_de_botopedia.

‘Wikipedia:Votaciones/2015/Modificación a la política de bots’. 2019. Wikipedia ES. 26 June 2019. https://es.wikipedia.org/wiki/Wikipedia:Votaciones/2015/Modificaci%C3%B3n_a_la_pol%C3%ADtica_de_bots.

‘Wikipedia:Wikipedians’. 2021. Wikipedia. 4 July 2021. https://en.wikipedia.org/wiki/Wikipedia:Wikipedians.

## 8.2 Secondary sources


Alegria, Iñaki, Unai Cabezon, Unai Fernandez de Betoño, Gorka Labaka, Aingeru Mayor, Kepa Sarasola, and Arkaitz Zubiaga. 2013. 'Reciprocal Enrichment Between Basque Wikipedia and Machine Translation'. In *The People's Web Meets NLP*, edited by Iryna Gurevych and Jungi Kim, 101–18. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35085-6_4.

Alonso, Elisa. 2015. 'Analysing the Use and Perception of Wikipedia in the Professional Context of Translation'. *The Journal of Specialised Translation* 23: 89–116.

Al-Shehari, Khaled. 2017. 'Collaborative Learning: Trainee Translators Tasked to Translate Wikipedia Entries from English into Arabic'. *The Interpreter and Translator Trainer* 11 (4): 357–72. https://doi.org/10.1080/1750399X.2017.1359755.

Aristotle. 2001. *Nicomachean Ethics*. Blacksburg, VA: Virginia Tech.

Ayers, Phoebe, Charles Matthews, and Ben Yates. 2008. *How Wikipedia Works : And How You Can Be a Part of It*. San Francisco, CA: No Starch Press, Incorporated.

Bilić, Paško. 2015. ''Searching for a Centre That Holds' in the Network Society: Social Construction of Knowledge on, and with, English Wikipedia'. *New Media & Society* 17 (8): 1258–76. https://doi.org/10.1177/1461444814522953.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge Studies in Social Anthropology ; 16. Cambridge ; New York: Cambridge University Press.

Bryman, Alan. 2016. *Social Research Methods*. Fifth Edition. Oxford ; New York: Oxford University Press.

Buch, Anders, and Theodore R. Schatzki, eds. 2019. *Questions of Practice in Philosophy and Social Theory*. Routledge Studies in Contemporary Philosophy. New York, NY: Routledge.

Callahan, Ewa S., and Susan C. Herring. 2011. 'Cultural Bias in Wikipedia Content on Famous Persons'. *Journal of the American Society for Information Science and Technology* 62 (10): 1899–1915. https://doi.org/10.1002/asi.21577.

Cámara de la Fuente, Lidia. 2015. 'Motivation to collaboration in TED Open Translation Project'. *International Journal of Web Based Communities* 11 (2): 210–229.

Chandra, Anita, and Abyayananda Maiti. 2018. 'Modeling New and Old Editors' Behaviors in Different Languages of Wikipedia'. In *Web Information Systems Engineering – WISE 2018*, edited by Hakim Hacid, Wojciech Cellary, Hua Wang, Hye-Young Paik, and Rui Zhou, 11234:438–53. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02925-8_31.

Clément, Maxime, and Matthieu J. Guitton. 2015. 'Interacting with Bots Online: Users' Reactions to Actions of Automated Programs in Wikipedia'. *Computers in Human Behavior* 50: 66–75. https://doi.org/10.1016/j.chb.2015.03.078.

Coutard, Olivier, and Elizabeth Shove. 2019. 'Infrastructures, Practices and the Dynamics of Demand'. In *Infrastructures in Practice: The Dynamics of Demand in Networked Societies*, 10–24. New York, NY: Routledge.

Dijk, Ziko van. 2009. 'Wikipedia and Lesser-Resourced Languages'. *Language Problems and Language Planning* 33 (3): 234–50. https://doi.org/10.1075/lplp.33.3.03van.

Ensslin, Astrid. 2011. ‘What an Un-Wiki Way of Doing Things: Wikipedia’s Multilingual Policy and Metalinguistic Practice’. *Journal of Language and Politics* 10 (4): 535–61. https://doi.org/10.1075/jlp.10.4.04ens.

Eysenbach, Gunther, and James E. Till. 2001. ‘Ethical Issues in Qualitative Research on Internet Communities’. *BMJ* 323 (7321): 1103–5. https://doi.org/10.1136/bmj.323.7321.1103.

Farnsworth, Valerie, Irene Kleanthous, and Etienne Wenger-Trayner. 2016. ‘Communities of Practice as a Social Theory of Learning: A Conversation with Etienne Wenger’. *British Journal of Educational Studies* 64 (2): 139–60. https://doi.org/10.1080/00071005.2015.1133799.

Ferschke, Oliver, Iryna Gurevych, and Yevgen Chebotar. 2012. ‘Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages’. In , 777–86. Avignon.

Gallus, Jana, and Sudeep Bhatia. 2020. ‘Gender, Power and Emotions in the Collaborative Production of Knowledge: A Large-Scale Analysis of Wikipedia Editor Conversations’. *Organizational Behavior and Human Decision Processes* 160: 115–30. https://doi.org/10.1016/j.obhdp.2020.02.003.

Geiger, R. Stuart. 2011. ‘The Lives of Bots’, Computers and Society, . https://arxiv.org/abs/1810.09590.

Geiger, R. Stuart, and Aaron Halfaker. 2017. ‘Operationalizing Conflict and Cooperation between Automated Software Agents in Wikipedia: A Replication and Expansion of “Even Good Bots Fight”’. *Proceedings of the ACM on Human-Computer Interaction* 1 (CSCW): 1–33. https://doi.org/10.1145/3134684.

Gherardi, Silvia. 2017. 'Sociomateriality in Posthuman Practice Theory'. In *The Nexus of Practices: Connections, Constellations, Practitioners*, edited by Allison Hui, Theodore Schatzki, and Elizabeth Shove, 38–51. London: Routledge.

Gillmeister, Heiner. 1997. *Tennis: A Cultural History*. London ; Washington [D.C.]: Leicester University Press.

Goldspink, Christopher. 2010. 'Normative Behavior in Wikipedia'. *Information, Communication & Society* 13 (5): 652–73. https://doi.org/10.1080/13691180903214523.

Góngora-Goloubintseff, José Gustavo. 2020. 'The Falklands/Malvinas War Taken to the Wikipedia Realm: A Multimodal Discourse Analysis of Cross-Lingual Violations of the Neutral Point of View'. *Palgrave Communications* 6 (1). https://doi.org/10.1057/s41599-020-0435-2.

Grasso, John. 2011. *Historical Dictionary of Tennis*. Historical Dictionaries of Sports. Lanham, Md: Scarecrow Press.

Gredel, Eva. 2017. 'Digital Discourse Analysis and Wikipedia: Bridging the Gap between Foucauldian Discourse Analysis and Digital Conversation Analysis'. *Journal of Pragmatics* 115: 99–114. https://doi.org/10.1016/j.pragma.2017.02.010.

Halfaker, Aaron, and John Riedl. 2012. 'Bots and Cyborgs: Wikipedia's Immune System'. *Computer* 45 (3): 79–82. https://doi.org/10.1109/MC.2012.82.

Hara, Noriko, Pnina Shachaf, and Khe Foon Hew. 2010. 'Cross-Cultural Analysis of the Wikipedia Community'. *Journal of the American Society for Information Science and Technology* 61 (10): 2097–2108. https://doi.org/10.1002/asi.21373.

Hautasaari, Ari, and Toru Ishida. 2012. 'Analysis of Discussion Contributions in Translated Wikipedia Articles'. In *Proceedings of the 4th International Conference on Intercultural Collaboration - ICIC '12*, 57. Bengaluru, India: ACM Press. https://doi.org/10.1145/2160881.2160891.

Hautasaari, Ari, Toshiyuki Takasaki, Takao Nakaguchi, Jun Koyama, Yohei Murakami, and Toru Ishida. 2011. 'Multi-Language Discussion Platform for Wikipedia Translation'. In *The Language Grid*, edited by Toru Ishida, 231–44. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21178-2_15.

Heidegger, Martin. 1929. *Being and Time*. Albany: SUNY Press.

Hui, Allison. 2017. 'Variation and the Intersection of Practices'. In *The Nexus of Practices: Connections, Constellations, Practitioners*, edited by Allison Hui, Theodore Schatzki, and Elizabeth Shove, First, 52–67. London: Routledge.

Jervell, Ellen Emmerentze. 2014. 'For This Author, 10,000 Wikipedia Articles Is a Good Day's Work'. *The Wall Street Journal*, 13 July 2014. https://www.wsj.com/articles/for-this-author-10-000-wikipedia-articles-is-a-good-days-work-1405305001.

Jiménez-Crespo, Miguel A. 2017. *Crowdsourcing and Online Collaborative Translations: Expanding the Limits of Translation Studies*. Benjamins Translation Library 131. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Jones, Henry. 2017. 'Multilingual Knowledge Production and Dissemination in Wikipedia : A Spatial Narrative Analysis of the Collaborative Construction of

City-Related Articles within the User-Generated Encyclopaedia'. Manchester: University of Manchester.

———. 2018a. 'Wikipedia as a Translation Zone: A Heterotopic Analysis of the Online Encyclopedia and Its Collaborative Volunteer Translator Community'. *Target* 31 (1): 77–97. https://doi.org/10.1075/target.18062.jon.

———. 2018b. 'Wikipedia, Translation, and the Collaborative Production of Spatial Knowledge'. *Alif: Journal of Comparative Poetics*, no. 38.

Jupp, Victor, ed. 2006. *The SAGE Dictionary of Social Research Methods*. London ; Thousand Oaks, Calif: SAGE Publications.

Khazraie, Marzieh, and Hossein Talebzadeh. 2020. '"Wikipedia Does NOT Tolerate Your Babbling!": Impoliteness-Induced Conflict (Resolution) in a Polylogal Collaborative Online Community of Practice'. *Journal of Pragmatics* 163: 46–65. https://doi.org/10.1016/j.pragma.2020.03.009.

King, Nigel, Christine Horrocks, and Joanna Brooks. 2019. *Interviews in Qualitative Research*. Second. Los Angeles: SAGE.

Klapper, Helge, and Markus Reitzig. 2018. 'On the Effects of Authority on Peer Motivation: Learning from Wikipedia'. *Strategic Management Journal* 39 (8): 2178–2203. https://doi.org/10.1002/smj.2909.

Kozinets, Robert V. 2002. 'The Field behind the Screen: Using Netnography for Marketing Research in Online Communities'. *Journal of Marketing Research* 39 (1): 61–72. https://doi.org/10.1509/jmkr.39.1.61.18935.

Kvale, Steinar, and Svend Brinkmann. 2009. *InterViews: Learning the Craft of Qualitative Research Interviewing*. 2nd ed. Los Angeles: Sage Publications.

Laat, Paul B. de. 2015. 'The Use of Software Tools and Autonomous Bots against Vandalism: Eroding Wikipedia's Moral Order?' *Ethics and Information Technology* 17 (3): 175–88. https://doi.org/10.1007/s10676-015-9366-9.

Lanamäki, Arto, and Juho Lindman. 2018. 'Latent Groups in Online Communities: A Longitudinal Study in Wikipedia'. *Computer Supported Cooperative Work (CSCW)* 27 (1): 77–106. https://doi.org/10.1007/s10606-017-9295-8.

Langer, Roy, and Suzanne C. Beckman. 2005. 'Sensitive Research Topics: Netnography Revisited'. Edited by Richard Elliott. *Qualitative Market Research: An International Journal* 8 (2): 189–203. https://doi.org/10.1108/13522750510592454.

Laniado, David, Riccardo Tasso, Yana Volkovich, and Andreas Kaltenbrunner. 2011. 'When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages'. In , 177–184.

Latour, Bruno. 1991. 'The Impact of Science Studies on Political Philosophy'. *Science, Technology, & Human Values* 16 (1): 3–19. https://doi.org/10.1177/016224399101600101.

Lave, Jean, and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Learning in Doing. Cambridge [England] ; New York: Cambridge University Press.

Laxström, Niklas, Pau Giner, and Santhosh Thottingal. 2015. 'Content Translation: Computer-Assisted Translation Tool for Wikipedia Articles'. *Computation and Language* EAMT 2015. https://arxiv.org/abs/1506.01914.

Leung, Lawrence. 2015. 'Validity, Reliability, and Generalizability in Qualitative Research'. *Journal of Family Medicine and Primary Care* 4 (3): 324. https://doi.org/10.4103/2249-4863.161306.

Lewoniewski, Włodzimierz. 2019. 'Measures for Quality Assessment of Articles and Infoboxes in Multilingual Wikipedia'. In *Business Information Systems Workshops*, edited by Witold Abramowicz and Adrian Paschke, 339:619–33. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-04849-5_53.

Lewoniewski, Włodzimierz, Krzysztof Węcel, and Witold Abramowicz. 2018. 'Determining Quality of Articles in Polish Wikipedia Based on Linguistic Features'. In *Information and Software Technologies*, edited by Robertas Damaševičius and Giedrė Vasiljevienė, 920:546–58. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-99972-2_45.

Lih, Andrew. 2009. *The Wikipedia Revolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia*. London ; New York: Aurum.

Littau, Karin. 2016. 'Translation and the Materialities of Communication'. *Translation Studies* 9 (1): 82–96. https://doi.org/10.1080/14781700.2015.1063449.

———. 2017. 'Response by Littau to the Responses to "Translation and the Materialities of Communication"'. *Translation Studies* 10 (1): 97–101. https://doi.org/10.1080/14781700.2016.1243486.

Lynch, Julianne, Julie Rowlands, Trevor Gale, and Andrew Skourdoumbis, eds. 2017. *Practice Theory and Education: Diffractive Readings in Professional Practice*. London ; New York, NY: Routledge is an imprint of the Taylor & Francis Group, an Informa Business.

Maller, Cecily Jane. 2015. 'Understanding Health through Social Practices: Performance and Materiality in Everyday Life'. *Sociology of Health & Illness* 37 (1): 52–66. https://doi.org/10.1111/1467-9566.12178.

Mann, Steve. 2016. *The Research Interview: Reflective Practice and Reflexivity in Research Processes*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave Macmillan.

Marx, Karl. 1962. *Tesis sobre Feuerbach*. Editora Política. Moscow, Havana. https://www.filosofia.org/cla/ome/45tes_b.htm.

Marx, Karl, and Friedrich Engels. 1867. *Das Kapital: Kritik Der Politischen Ökonomie*. Hamburg: Meissner.

Mattus, Maria. 2014. 'The Anyone-Can-Edit Syndrome: Intercreation Stories of Three Featured Articles on Wikipedia'. *Nordicom Review* 35 (s1): 189–204. https://doi.org/10.2478/nor-2014-0112.

McDonough Dolmaya, Julie. 2012. 'Analyzing the Crowdsourcing Model and Its Impact on Public Perceptions of Translation'. *The Translator* 18 (2): 167–91. https://doi.org/10.1080/13556509.2012.10799507.

———. 2015. 'Revision History: Translation Trends in Wikipedia'. *Translation Studies* 8 (1): 16–34. https://doi.org/10.1080/14781700.2014.943279.

———. 2017. ‘Expanding the Sum of All Human Knowledge: Wikipedia, Translation and Linguistic Justice’. *The Translator* 23 (2): 143–57. https://doi.org/10.1080/13556509.2017.1321519.

McDonough Dolmaya, Julie, and María Sánchez-Ramos. 2019. ‘Characterizing Online Social Translation’. *Translation Studies* 12 (2): 129–38. https://doi.org/10.1080/14781700.2019.1697736.

Melo, Gerard de, and Gerhard Weikum. 2014. ‘Taxonomic Data Integration from Multilingual Wikipedia Editions’. *Knowledge and Information Systems* 39 (1): 1–39. https://doi.org/10.1007/s10115-012-0597-3.

Morgan, Jonathan T., and Mark Zachry. 2010. ‘Negotiating with Angry Mastodons: The Wikipedia Policy Environment as Genre Ecology’. In *Proceedings of the 16th ACM International Conference on Supporting Group Work - GROUP ’10*, 165. Sanibel Island, Florida, USA: ACM Press. https://doi.org/10.1145/1880071.1880098.

Neather, Robert. 2012. ‘“Non-Expert” Translators in a Professional Community: Identity, Anxiety and Perceptions of Translator Expertise in the Chinese Museum Community’. *The Translator* 18 (2): 245–268. https://doi.org/10.1080/13556509.2012.10799510.

Nemoto, Keiichi, and Peter A. Gloor. 2011. ‘Analyzing Cultural Differences in Collaborative Innovation Networks by Analyzing Editing Behavior in Different-Language Wikipedias’. *Procedia - Social and Behavioral Sciences* 26: 180–90. https://doi.org/10.1016/j.sbspro.2011.10.574.

Nicolini, Davide. 2011. ‘Practice as the Site of Knowing: Insights from the Field of Telemedicine’. *Organization Science* 22 (3): 602–20. https://doi.org/10.1287/orsc.1100.0556.

———. 2012. *Practice Theory, Work, and Organization : An Introduction*. Oxford: Oxford University Press.

O’Brien, Sharon, and Reinhard Schäler. 2010. ‘Next generation translation and localization: Users are taking charge’. Paper presented at *Translating and the Computer 32, London, November 18–19*’. http://doras.dcu.ie/16695/1/Paper_6.pdf.

Oeberst, Aileen, Ina von der Beck, Ulrike Cress, and Steffen Nestler. 2020. ‘Wikipedia Outperforms Individuals When It Comes to Hindsight Bias’. *Psychological Research* 84 (6): 1517–27. https://doi.org/10.1007/s00426-019-01165-7.

O’Hagan, Minako. 2016. ‘Massively Open Translation: Unpacking the Relationship Between Technology and Translation in the 21st Century’. *International Journal of Communication (Online)*.

Olohan, Maeve. 2012. ‘Volunteer Translation and Altruism in the Context of a Nineteenth-Century Scientific Journal’. *The Translator* 18 (2): 193–215. https://doi.org/10.1080/13556509.2012.10799508.

———. 2014. ‘History of Science and History of Translation: Disciplinary Commensurability?’. *The Translator* 20 (1): 9–25. https://doi.org/10.1080/13556509.2014.899091.

———. 2014. ‘Why do you translate? Motivation to Volunteer and TED Translation’. *Translation Studies* 7 (1): 17–33. https://doi.org/10.1080/14781700.2013.781952.

———. 2017. ‘Knowing in Translation Practice: A Practice-Theoretical Perspective’. *Translation Spaces* 6 (1): 159–80. https://doi.org/10.1075/ts.6.1.08olo.

———. 2021. *Translation and Practice Theory*. 1st ed. Translation Theories Explored. London ; New York: Routledge.

O’Sullivan, Dan. 2009. *Wikipedia: A New Community of Practice?* Farnham, England ; Burlington, VT: Ashgate.

Park, Chan Young, Xinru Yan, Anjalie Field, and Yulia Tsvetkov. 2020. ‘Multilingual Contextual Affective Analysis of LGBT People Portrayals in Wikipedia’. *ICWSM 2021*, Computation and Language.

Pee, Loo Geok. 2018. ‘Community’s Knowledge Need and Knowledge Sharing in Wikipedia’. *Journal of Knowledge Management* 22 (4): 912–30. https://doi.org/10.1108/JKM-09-2017-0412.

Pérez-González, Luis, and Şebnem Susam-Saraeva. 2012. ‘Non-Professionals Translating and Interpreting: Participatory and Engaged Perspectives’. *The Translator* 18 (2): 149–65. https://doi.org/10.1080/13556509.2012.10799506.

Pfeil, Ulrike, Panayiotis Zaphiris, and Chee Siang Ang. 2006. ‘Cultural Differences in Collaborative Authoring of Wikipedia’. *Journal of Computer-Mediated Communication* 12 (1): 88–113. https://doi.org/10.1111/j.1083-6101.2006.00316.x.

Reckwitz, Andreas. 2002. 'Toward a Theory of Social Practices: A Development in Culturalist Theorizing'. *European Journal of Social Theory* 5 (2): 243–63. https://doi.org/10.1177/13684310222225432.

Roberts, Lynne D. 2015. 'Ethical Issues in Conducting Qualitative Research in Online Communities'. *Qualitative Research in Psychology* 12 (3): 314–25. https://doi.org/10.1080/14780887.2015.1008909.

Saldanha, Gabriela, and Sharon O'Brien. 2013. *Research Methodologies in Translation Studies*. Manchester, UK: St. Jerome Publishing.

Saldaña, Johnny. 2021. *The Coding Manual for Qualitative Researchers*. 4th ed. Thousand Oaks, California: SAGE Publications.

Salmons, Janet. 2016. *Doing Qualitative Research Online*. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. https://doi.org/10.4135/9781473921955.

Samoilenko, Anna, Fariba Karimi, Daniel Edler, Jérôme Kunegis, and Markus Strohmaier. 2016. 'Erratum to: Linguistic Neighbourhoods: Explaining Cultural Borders on Wikipedia through Multilingual Co-Editing Activity'. *EPJ Data Science* 5 (1): 15. https://doi.org/10.1140/epjds/s13688-016-0076-2.

Schatzki, Theodore. 2003. 'A New Societist Social Ontology'. *Philosophy of the Social Sciences* 33 (2): 174–202. https://doi.org/10.1177/0048393103033002002.

———.2010. 'Materiality and Social Life'. *Nature and Culture* 5 (2): 123–49. https://doi.org/10.3167/nc.2010.050202.

———. 2018. ‘On Practice Theory, or What’s Practices Got to Do (Got to Do) with It?’ In *Education in an Era of Schooling*, edited by Christine Edwards-Groves, Peter Grootenboer, and Jane Wilkinson, 151–65. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-2053-8_11.

Schmidt, Robert. 2017. ‘Reflexive Knowledge in Practices’. In *The Nexus of Practices: Connections, Constellations, Practitioners*, edited by Allison Hui, Theodore Schatzki, and Elizabeth Shove, First, 141–54. Routledge.

Schneider, Jodi, Alexandre Passant, and John G. Breslin. 2010. ‘A Content Analysis: How Wikipedia Talk Pages Are Used’. In , 1–7. Raleigh, NC. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.3779&rep=rep1&type=pdf.

Shane-Simpson, Christina, and Kristen Gillespie-Lynch. 2017. ‘Examining Potential Mechanisms Underlying the Wikipedia Gender Gap through a Collaborative Editing Task’. *Computers in Human Behavior* 66: 312–28. https://doi.org/10.1016/j.chb.2016.09.043.

Shi, Feng, Misha Teplitskiy, Eamon Duede, and James A. Evans. 2019. ‘The Wisdom of Polarized Crowds’. *Nature Human Behaviour* 3 (4): 329–36. https://doi.org/10.1038/s41562-019-0541-6.

Shove, Elizabeth. 2017. ‘Matters of Practice’. In *The Nexus of Practices: Connections, Constellations, Practitioners*, edited by Allison Hui, Theodore Schatzki, and Elizabeth Shove, First, 155–68. London: Routledge.

Shove, Elizabeth, Mika Pantzar, and Matt Watson. 2012. *The Dynamics of Social Practice: Everyday Life and How It Changes*. Los Angeles: SAGE.

Shove, Elizabeth, and Frank Trentmann, eds. 2019. *Infrastructures in Practice: The Dynamics of Demand in Networked Societies*. Abingdon, Oxon ; New York, NY: Routledge.

Shuttleworth, Mark. 2017. 'Locating Foci of Translation on Wikipedia: Some Methodological Proposals'. *Translation Spaces* 6 (2): 310–32. https://doi.org/10.1075/ts.6.2.07shu.

———. 2018. 'Translation and the Production of Knowledge in Wikipedia: Chronicling the Assassination of Boris Nemtsov'. *Alif: Journal of Comparative Poetics*, no. 38.

*The Local*. 2014. 'Swedish Wiki Vet Sets New Content Record', 18 July 2014. https://www.thelocal.se/20140718/swedish-wiki-vet-sets-new-record/.

Torres-Simón, Ester. 2019. 'The Concept of Translation in Wikipedia'. *Translation Studies* 12 (3): 273–87. https://doi.org/10.1080/14781700.2018.1534605.

Townsend, Stephen, Gary Osmond, and Murray G. Phillips. 2013. 'Wicked Wikipedia? Communities of Practice, the Production of Knowledge and Australian Sport History'. *The International Journal of the History of Sport* 30 (5): 545–59. https://doi.org/10.1080/09523367.2013.767239.

Tsvetkova, Milena, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. 2017. 'Even Good Bots Fight: The Case of Wikipedia'. Edited by Sergio Gómez. *PLOS ONE* 12 (2): e0171774. https://doi.org/10.1371/journal.pone.0171774.

'University Ethical Approval'. 2021. University of Manchester. 2021. https://www.manchester.ac.uk/research/environment/governance/ethics/approval/.

Usman, Berto, and Yennita Yennita. 2018. ‘A Glimpse into the Virtual Community of Practice (CoP): Knowledge Sharing in the Wikipedia Community’. *Asian Journal of Business and Accounting* 11 (2): 249–76. https://doi.org/10.22452/ajba.vol11no2.8.

Viégas, Fernanda B., Martin Wattenberg, and Matthew M. McKeon. 2007. ‘The Hidden Order of Wikipedia’. In *Online Communities and Social Computing*, edited by Douglas Schuler, 4564:445–54. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73257-0_49.

Vrandečić, Denny. 2020. ‘Architecture for a Multilingual Wikipedia’. *Computers and Society*, August. https://arxiv.org/abs/2004.04733.

Warde, Alan. 2005. ‘Consumption and Theories of Practice’. *Journal of Consumer Culture* 5 (2): 131–53. https://doi.org/10.1177/1469540505053090.

———. 2016. *The Practice of Eating*. Malden, MA: Polity Press.

Wenger, Etienne. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Learning in Doing. Cambridge, U.K. ; New York, N.Y: Cambridge University Press.

Wenger, Etienne, Richard A. McDermott, and William Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Business & Economics. Boston: Harvard Business Press.

Wenger, Etienne, Nancy White, and John D. Smith. 2009. *Digital Habitats : Stewarding Technology for Communities*. Portland, OR: CPsquare.

Wittgenstein, Ludwig. 1981. *Zettel*. Second. Oxford: Blackwell.

Yli-Kauhaluoma, Sari, and Mika Pantzar. 2016. ‘Intricacies of Back-Office: Forward- and Backward-Looking Recipes as Modelling Tools in Administration’. *Journal of Organizational Ethnography* 5 (2): 167–83. https://doi.org/10.1108/JOE-01-2016-0005.

Young, Amber G., Ariel D. Wigdor, and Gerald C. Kane. 2020. ‘The Gender Bias Tug-of-War in a Co-Creation Community: Core-Periphery Tension on Wikipedia’. *Journal of Management Information Systems* 37 (4): 1047–72. https://doi.org/10.1080/07421222.2020.1831773.

Yu, Chuan. 2019. ‘Negotiating identity roles during the process of online collaborative translation: An ethnographic approach’. *Translation Studies* 12 (2): 231-252. https://doi.org/10.1080/14781700.2019.1692062.

# Appendix I Participant Information Sheet

## TRANSLATION IN WIKIPEDIA: A PRAXEOLOGICAL STUDY OF NORMATIVITY, NEGOTIATION AND AUTOMATION ACROSS FOUR LANGUAGE COMMUNITIES

**Participant Information Sheet (PIS)**

This PIS should be read in conjunction with The University privacy notice.

You are being invited to take part in a research study as part of a student PhD project that aims to investigate how translation practices are regulated, negotiated and performed across four language communities of Wikipedia. Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

**About the research**

**Who will conduct the research?**

José Gustavo Góngora-Goloubintseff

PhD student

University of Manchester

School of Arts, Languages and Cultures

Samuel Alexander Building

M13 9PL

**What is the purpose of the research?**

The aim of this study is to investigate how translation practices are regulated, negotiated and performed across the Dutch, French, Spanish and Swedish language communities of Wikipedia. The investigation also seeks to gain insight into how experienced editors of the four communities use programs such as the Content Translation Tool (CX) and software robots (bots).

**Will the outcomes of the research be published?**

The thesis will be stored in PDF format on a public online repository once it has been successfully defended. In addition, publishable works generated from the PhD will also be available at your request.

**Who has reviewed the research project?**

This project has been reviewed and approved by the researcher's main supervisor at the University of Manchester.

**Who is funding the research project?**

This project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under the Doctoral Training Partnership scheme (DTP). Project reference: 2068270.

**What would my involvement be?**

**Why have I been chosen?**

You have been chosen because you have prior experience as a translator in Wikipedia. You have been a registered Wikipedia editor for at least three years and your record of contributions indicate that you have been actively involved in the online encyclopaedia during that time. There will be 16 participants taking part in this study, four per language community.

**What would I be asked to do if I took part?**

You will be invited to engage in a **one-to-one online interview**, which will be held using Zoom. During the interview, you will be asked to reflect on your experience as a translator in Wikipedia. Specifically, the researcher will ask you about the following themes:

- Your experience as a Wikipedia translator
- The type of articles you have translated
- Wikipedia policies or guidelines you adhere to when you translate
- If applicable, the type of automated tools you have used for your translations
- If applicable, your experience using such tools
- Your opinion on and, if applicable, your experience with, the use of bots for content-creation purposes
- Your opinion on how automated tools and Wikidata have contributed to changes in translation practices within your Wikipedia community

The interview is set to last from 30 to 60 minutes.

No sensitive questions will be asked during the interview and the participant will neither be required nor expected to disclose personal information such as their real name or location. However, information such as age group, profession and level of education will be collected in a separate document, subject to the participant's approval, prior to the interview.

The interview will be audio-recorded and will be conducted in English or Spanish if required.

**Will I be compensated for taking part?**

Unfortunately, no monetary incentives will be provided for your participation in this research project.

**What happens if I do not want to take part or if I change my mind?**

It is up to you to decide whether to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time without giving a reason and without detriment to yourself. However, it will not be possible to remove your data from the project once it has been anonymised and forms part of the dataset, as we will not be able to identify your specific data. This does not affect your data protection rights.

At the agreed date of the interview, you may freely decline if you are not feeling well by any reason. However, since recorded audios are essential to the research, you will be asked to reschedule the interview. If you do not wish to be interviewed any further, then you can withdraw from the research altogether.

In addition, you should feel comfortable at all times during the interview and you are free to stop the recording at any time should this not be the case.

**Data Protection and Confidentiality**

**What will happen to my personal information?**

In order to undertake the research project, we will need to collect the following personal information/data about you:

- Audio recordings regarding your experience as a translator in Wikipedia
- Age group, profession, and level of education (to be asked in a separate Word document before the interview)

In the case of interview, this will be conducted on a one-to-one basis via Zoom or similar software. The interview will be voice recorded only and it will be used solely for the purposes of this research as well as in any potential publications that emerge from the findings of this work. Data generated will not be reused.

The researcher and his supervisor will have access to this information, as well as the transcribers, but in the case of the latter, data will be encrypted to guarantee confidentiality. Other participants will not have access to what you say, nor will you have access to their interviews.

We are collecting and storing this personal information in accordance with the General Data Protection Regulation (GDPR) and Data Protection Act 2018, which legislate to protect your personal information. The legal basis upon which we are using your personal information is "public interest task" and "for research purposes" if personal information is collected. For more information about the way we process your personal information and comply with data protection law, please see our Privacy Notice for Research Participants.

The University of Manchester, as Data Controller for this project, takes responsibility for the protection of the personal information that this study is collecting about you. In order to comply with the legal obligations to protect your personal data, the University has safeguards in place such as policies and procedures. All researchers are appropriately trained and your data will be looked after in the following way:

The researcher and his supervisor at the University of Manchester will have access to your personal identifiable information, that is data which could identify you, but these will be anonymised as soon as the interview session finishes. However, your consent form and contact details will be retained as long as the research is in progress and will be destroyed as soon as the PhD has been successfully completed. Your data will be stored in the university Research Data Storage (RDS), which can only be accessed with credentials known to the researcher.

Transcribers will be asked to sign a confidentiality agreement and then will be granted access to your encrypted data. This means they will not be able to track you nor identify you. As stated above, other participants in the project will not have access to your data nor to what you have stated in your interview. This also means you will not have access to what they say or share with the researcher. All these data will be destroyed once the study finishes.

You have a number of rights under data protection law regarding your personal information. For example, you can request a copy of the information we hold about you, including audio recordings. This is known as a Subject Access Request. If you would like to know more about your different rights, please consult our privacy notice for research and if you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, University of Manchester, Oxford Road, M13 9PL. at the University and we will guide you through the process of exercising your rights.

You also have a right to complain to the Information Commissioner's Office, Tel 0303 123 1113.

**Will my participation in the study be confidential?**

Your participation in the study will be kept confidential to the researcher and those with access to your personal information as listed above.

The transcription will be done by proficient English speakers who will be hired by the researcher. Transcribers will be required to sign a confidentiality agreement covering their specific involvement in this PhD project. The audio files will be password protected (encrypted) and will be shared using links to Dropbox folders. Your personal data will be removed and replaced using a standard protocol given to the transcribers once they have signed the agreement. Transcribers will be requested to delete any copies they might have created while transcribing the file, and access to the audio file will be removed once they deliver the transcribed version of the files. Upon receiving the transcribed documents, these will also be stored in the RDS server while the PhD is in progress.

After the PhD has been successfully completed, data will be stored for up to six months in compliance with the procedures outlined above and then will be erased from the server.

**What is the duration of the research?**

The research will last 14 months between July 2020 and September 2021. You will be asked to participate in only one interview, which will last from 30 to 60 minutes. The interview will be scheduled at a time suitable to you.

**What if I have a complaint?**

**Minor complaints**

If you have a minor complaint, you need to contact the researcher(s) in the first instance.

Researcher:

JOSE GUSTAVO GONGORA-GOLOUBINTSEFF

jose.gongoragoloubintseff@postgrad.manchester.ac.uk

Supervisor:

DR MAEVE OLOHAN

maeve.olohan@manchester.ac.uk

Tel +44 (0) 161 275 3128

**Formal Complaints**

**If you wish to make a formal complaint or if you are not satisfied with the response you have gained from the researchers in the first instance then please contact**

The Research Governance and Integrity Manager, Research Office, Christie Building, University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing: research.complaints@manchester.ac.uk or by telephoning +44 (0) 161 275 2674.

**Contact Details**

If you have any queries about the study or if you are interested in taking part then please contact the researcher(s)**.**

**JOSE GUSTAVO GONGORA-GOLOUBINTSEFF**

jose.gongoragoloubintseff@postgrad.manchester.ac.uk

# Appendix II Participant Consent Form

**TRANSLATION IN WIKIPEDIA: A PRAXEOLOGICAL STUDY OF NORMATIVITY, NEGOTIATION AND AUTOMATION ACROSS FOUR LANGUAGE COMMUNITIES**

**Participant Consent Form**

If you are happy to participate, please complete and sign the consent form below.

| | Activities | Initials |
|---|---|---|
| 1 | I confirm that I have read the attached information sheet (**Version 1.0, Date 13/07/2020**) for the above study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily. | |
| 2 | I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to myself. I understand that it will not be possible to remove my data from the project once it has been anonymised and forms part of the data set.<br><br>I **agree** to take part on this basis. | |
| 3 | I agree to the **interviews** being **audio recorded**. | |
| 4 | I agree that any data collected may be published in anonymous form in **academic books, reports or journals.** | |
| 5 | I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study. | |
| 6 | I agree to take part in this study. | |

**Data Protection**

**The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.**

| _____ | _____ | _____ |
|---|---|---|
| Name of Participant | Signature | Date |
| _____ | _____ | _____ |
| Name of the person taking consent | Signature | Date |

There will be two copies of this document: 1 for the participant and 1 copy for the researcher (original).

# Appendix III Interview guide

| Themes | Questions |
|---|---|
| **Background** | 1. How long have you been a registered editor in Wikipedia? |
| | 2. Why did you decide to become a translator in Wikipedia? |
| | 3. How would you describe your experience as a translator in Wikipedia? |
| | 4. What kind of articles have you translated?<br>• Do you include the references of the original Wikipedia article in your translation? Why/why not? |
| | 5. What are the major difficulties you faced when you started to translate articles in Wikipedia? How did you overcome them? |
| | 6. What criteria did you follow when choosing the articles you wanted to translate? |
| **Policies and guidelines** | 7. In your view, what are the most critical Wikipedia policies and guidelines that translators should follow?<br>• Which ones do you tend to prioritise when you translate? |
| | 8. In your view, how important is it to credit the original Wikipedia article when you have posted your translation?<br>• How do you usually do that? |

| Themes | Questions |
|---|---|
| **Automation** | 9. What resources do you normally use when you translate? |
| | 10. What do you think about using bots to create Wikipedia articles?<br>• In your view, what kind of articles are more likely to be created by bots? |
| | 11. Have you ever used automatic translation tools? If so, how was your experience using them? Do you still use them? Why/why not?<br>• Have you used Wikipedia's Content Translation Tool (CX)? If so, could you tell me more about your experience using the tool? What were the pros and cons of using CX? |
| **Evolution of translation in Wikipedia** | 12. In your view, have there been important changes in the way people translate in your Wikipedia community over the last years?<br>• If so, could you tell me more about it?<br>• In your view, have automatic translation tools contributed to those changes? Why/why not? |
| | 13. How do you envisage the future of translation in Wikipedia?<br>• How would you describe the role of Wikidata in that scenario, i.e. infoboxes? |
| | 14. Is there anything else you would like to add or comment on? |

# Appendix IV Questionnaire

Thanks for taking part in this study. Below are a series of background questions you should fill out before the interview. If there is something you do not understand, feel free to contact the researcher at jose.gongoragoloubintseff@postgrad.manchester.ac.uk

Once you have finished, please send this document back to the researcher via email.

Please, write your full name here: _____

| | Questions | Answer |
|---|---|---|
| 1 | Do you have a university degree?<br>a. Yes<br>b. No | |
| 2 | If the answer to the previous question is 'Yes', what did you study? | |
| 3 | What is your age group?<br>a. 18-30<br>b. 31-40<br>c. 41-50<br>d. >50 | |
| 4 | Using Wikipedia Babel categories, how would you rate your knowledge of English?<br>1. Basic<br>2. Intermediate<br>3. Advanced<br>4. Near native<br>5. Professional<br>6. Native | |
| 5 | Are you fluent in another language (apart from English)?<br>a. Yes (please specify)<br>b. No | |
| 6 | Did you have experience in translation before joining Wikipedia?<br>a. Yes<br>b. No | |
| 7 | If the answer to the previous question is 'Yes', which of the following options best describes your situation?<br>a. Professional translator (including freelance)<br>b. Volunteer translator (non-paid)<br>c. Other (please specify) | |

**Data Protection**

**The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.**