Nonhuman language agents in online collaborative communities: Comparing Hebrew Wikipedia and Facebook translations

Carmel L. Vaisman [a,*], Illan Gonen [b], Yuval Pinter [c]

[a] Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv, Israel
[b] Columbia University, NY, NY, United States
[c] School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

ARTICLE INFO

ABSTRACT

Online mass collaborations act as unregulated superdiverse language spaces, however, grassroots policing may impose uniformity and reproduce hegemony. This study compared language policies in Hebrew Wikipedia and the Hebrew Facebook translation app. Hebrew Wikipedia designed a strict linguistic guide that promotes a neutral Hebrew register, rejecting both colloquial and high registers, enforced by an algorithm post factum. The Hebrew Facebook translators' community maintained a decentralized approach, lacking the affordances for hierarchies of expertise, focusing on the practicality of the language and the speed of project completion. The comparative design within the same speech community stressed the role of affordances as nonhuman language agents in the social process of language policy.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

New media offer a plethora of unregulated spaces for mass collectives governed by an ethos of participatory text production. Wiki projects, and especially Wikipedia, are key examples, and indeed many works have documented the emerging practices that facilitate and sustain such collaborations that result in a single (though dynamic) text (e.g. Meyer, 2010; Regale, 2010), with some attention given to language issues (Emigh and Herring, 2005; Ensslin, 2011).

Since 2008, platforms such as Facebook and Twitter have solicited non-expert volunteers for crowdsourced translation projects of their interfaces, thus facilitating additional collaborative collectives with a specific linguistic focus. To date only one study has addressed the emergence of language ideologies in one such project, namely, the Irish Facebook translation app (Lenihan, 2011, 2014). The focus in this study was on language purism and policing of discussions among the community of translators (Lenihan, 2011, 2014).

The unregulated spaces of new media are inherently polyvocal, constituting the superdiversity (Blommaert and Rampton, 2011) of language, culture, and identity. Thus, they offer compelling case studies for observing the emergence of bottom-up self-regulation norms as a social process. Accordingly, the research question of the present study was, do these norms reproduce those of traditional authorities or challenge them? Addressing this question is important to advancing our understanding of multilingualism online and of language policy facets, such as policing of normativity and heteroglossia.

This study compared two online collaborative collectives: Hebrew Wikipedia and the Hebrew Facebook translation app. Because the majority of Hebrew Wikipedia is in fact translated from English Wikipedia, we are for the most part comparing two translation projects. The use of the comparative design within the same speech community was expected to enhance the understanding of language policy as a dynamic discursive process, while stressing the role of digital contexts and affordances.

The Hebrew Wikipedia and Hebrew Facebook projects have similar starting points: open to all, voluntary, and nonhierarchical in theory. Because they act as unregulated media spaces regarding the use of Hebrew, we expected to see similar processes of negotiating language ideologies and practices; yet, they developed entirely different language policies. In our analysis of these differences, we highlight the important role of nonhuman actors (e.g., technical affordances, algorithms) and consider the implications of these differences for other contexts of mediated communication.

* Corresponding author.
E-mail addresses: Carmell@post.tau.ac.il (C.L. Vaisman), ig2276@columbia.edu (I. Gonen), uvp@gatech.edu (Y. Pinter).

1.1. From language policy to language policing

The emergence of new media has engendered developments in the theoretical framework of language policy. Language policy refers to the sum of beliefs, practices, planning, and management of languages operating within an environment (Spolsky, 2004). Traditionally, the main actor in language policy has been the nation-state, however, it shares the playfield with new actors that focus on accent and discourse, thus creating a market in which sharp distinctions between speaking right and speaking wrong are articulated (Blommaert, 2009). An example of such an actor could be business corporations imposing standardized formats of voice tone, vocabulary, and questions structure, thus styling the speech of their customer service workers and enforcing it through surveillance practices (Cameroon, 2000). However, the role of speech communities in creating language policy has been increasing too (Spolsky, 2009), as the interaction between "bottom-up" language practices of speakers and imposed "top-down" policies serves to transform language policy (Shohamy, 2006).

The study of media in language policy research is limited, with existing studies examining issues such as the role of the press in language ideology debates (Androutsoupolos, 2009). New media, however, are drawing increased research attention in recent years as sites of language struggle that provide spaces for the rise of new agents that challenge dichotomous notions of language policy (Blommaert, 2009), such as the dichotomy of top-down/bottom-up processes in a nonhierarchical self-regulated community (Lenihan, 2014). Specifically, using the case of the Irish Facebook translation app, it was found that both Facebook Inc. and translator community members acted in ways that could be interpreted as bottom-up or top-down, according to the situation (Lenihan, 2014).

As a result, mediated communication research arguably (Blommaert et al., 2009) calls for a conceptual shift in scholarship: from a focus on multilingualism and English language domination to heteroglossia (style, register, genre, and lexis within a given language), emphasizing the multiplicity of voices and agency over the monolithic view of a "language"; and from a rigid and static notion of language policy to a variety of normativity-policing practices.

By employing Foucault's notion of "police", Blommaert (2009) charted the shift of focus from "policy" to "policing"; that is, the production of order—normatively organized and policed conduct—that is infinitely detailed, regulated, and carried out by multiple agents in different media settings. The concept of policing stresses language policy as a social process that could be undertaken by anyone in contexts with no set source of authority (Bres and Belling, 2015). Research examining online language policies is mainly concerned with language tolerance and multilingualism (Danet and Herring, 2007; Ensslin, 2011), yet according to Blommaert (2009), registers, styles, and lexis are also policed.

Web environments offer a unique opportunity to follow natural discourse in polyvocal groups. Such environments are sometimes seen as unregulated spaces driven by a nonhierarchical collaborative and participatory ethos (Lankshear and Knobel, 2007), when in fact their emerging self-regulation practices often entail internal policing practices that turn them into highly regulated linguistic environments (Androutsoupolos, 2009). For instance, according to Bres and Belling (2015), Facebook group administrators act as new agents of language policy who assume authority and police speakers, resulting in the reassertion of language policies that hinder the cultivation of a democratic participatory community (p. 372).

Large-scale collaborative projects on social media platforms constitute the superdiversity (Blommaert and Rampton, 2011) of language, culture, and identity, but when their end result is a unified text, they may be prone to reproduce modernist ideologies that impose uniformity, homogeneity, and even traditional hegemony. Indeed, in a study on a translation of the Facebook interface, Lenihan (2011) noted that Facebook's translation project exemplifies "fake multilingualism" (Kelly-Holmes, 2005) because its default language is always English. Translations must be submitted from the original American English, and when corresponding with Facebook, Inc., even in relation to issues of translation, their demand is for English communication only. Similarly, Ensslin (2011) argued that Wikipedia is an example of "potential multilingualism" (Edwards, 1995), demonstrating that Anglo-conformity is inevitably inscribed in its user interface.

1.2. Hebrew language policy and practices in context

In the nineteenth century, following the nationalist movements in Europe and Zionist lobbying for a Jewish state, an informal spoken vernacular of Hebrew was developed in addition to the sacred and literary variety (Spolsky, 2014). Language policy was driven by ideology, using Hebrew as the mobilizing symbol of the Zionist movement and later the State of Israel, while rejecting all other home languages of immigrants, especially diasporic Jewish languages such as Yiddish and Ladino (Shohamy, 2006).

The Academy of the Hebrew Language (AHL) takes responsibility for the corpus planning of Hebrew (Shohamy, 2006). The AHL was legislated in 1953. Its plenum consists of 23 members and 15 academic advisors, including outstanding scholars of linguistics and Judaic/Bible studies, poets, writers, and translators (AHL, n.d.). The AHL considers its decisions binding for all written texts and formal speech and for all governmental agencies, including the Israel Broadcasting Authority, but does not police spontaneous speech (AHL, n.d.). The main purpose of the Academy is not to stamp out all non-Hebrew influences, as is clear from the very name of the institution, which includes a transliterated foreign word (*academia*) in its name. The decision not to invent a Hebrew word for academia was fiercely debated in the first years of the Israeli government and is well documented (AHL, n.d.).

Modern Hebrew had been influenced by various foreign languages and its grammar now shows many Indo-European rather than Semitic features (Spolsky, 2014). Despite the lexical enrichments of the AHL, Modern Hebrew vocabulary includes borrowed words from all the languages of the Jews, as well as from international vocabularies (Spolsky, 2014; Zuckerman, 2003). This has led to the argument that Modern Hebrew is so different from biblical Hebrew that it should be called "Israeli" rather than Hebrew (Zuckerman, 2003).

Since the 1980s, characteristics of spoken Hebrew and slang have gradually encroached on print journalism, and since the 1990s have been included in dictionaries, including the standard-bearer Even-Shoshan Dictionary, beginning with its 2003 edition (Fruchtman, 2006). The specific problem of English influence on Modern Hebrew has been addressed by the AHL and many Israeli linguists (e.g., Muchnik, 1994; Schwarzwald, 1998), including giving Hebrew contemporary hybrid names such as Hebrish and Engrew.

Although English was one of the official languages in pre-state Israel during the British Mandate (1920–1948), its wide impact on Hebrew dates back only to the 1980s and is due to the globalization of culture and communication and incursions from American culture into Israeli society (Muchnik, 1994). The presence of English words and cultural markers in both spoken and written language has become so pervasive that many Hebrew speakers mistake them for original Hebrew slang; for instance, "cold feet" or "made my day" (Fisherman, 2006).

Existing evidence on the heteroglossic aspects of Hebrew language policing is limited and indirect. Some scholars have provided anecdotal evidence on imposition attempts of proper grammar on