EXTRALINGUISTIC ARGUMENTS IN 21ST CENTURY LANGUAGE PLANNING DISCOURSE: A “SUPERDICTIONARY” BETWEEN LANGUAGE STANDARDIZATION FROM ABOVE AND BELOW

Kadri Koreinik
University of Tartu, EE
kadri.koreinik@ut.ee

Abstract. Language standardization has historically been a critical area of inquiry in language policy and planning (LPP) research. This is a political matter, which contributes to “more (and hierarchical) heterogeneity” rather than linguistic homogeneity (Gal 2006: 171). The paper empirically explores extralinguistic arguments, which are used by language professionals (planners, academics, educators) in mainstream media discourse. This public discourse is initiated by the launch of an Estonian “superdictionary” in 2019 (see Tavast et al. 2020), and its public reception. By using Critical Discourse Analysis (CDA) as a method, the paper also provides insight into the discursive construction of language as such and (Standard) Estonian by different LPP actors. Above all, it aims to understand the issues of power and authority in language standardization. The discourse illustrates the paradigmatic change in standardization and lexicography: from including selected language samples to the acceptance of non-elite language variants and varieties. This change has generated a polarization of stance among language professionals, and similar discursive moves, e.g., references to the past and future dangers, metaphors and other comparisons are used.

Keywords: argumentation, language standardization, language ideologies, discourse, language planning, Estonian


1. Introduction: why this paper now?

Metalinguistic texts such as dictionaries, grammars, guides, manuals, columns, blogs, and shared social media content settle and record “what is correct and/or possible in language and what is not”, i.e., most of these have a codifying function according to the models of language standardization (McLelland 2021: 263; Haugen 1971, 1983). There are different types of codification: standard-creating,
modernizing, archaizing, and standard-descriptive, which mostly end up forming hybrid types (Zgusta 2006). While dictionaries are central to the creation, preservation, and modernization of standard languages, the relationship between lexicography and linguistics – let alone, linguistics and language cultivation – has never been simple.

Lexicography is a human practice which lacks its own theory, but one of its products – the dictionary – should aim to balance data, the user, and access to data (Bogaards 2010: 317). It “requires recourse to theories of data (linguistics), of the user (psychology, psycholinguistics, sociology), and of access to data (IT)” (Adamska-Sałaciak 2019: 13). In the popular imagination, national language planning agencies, also responsible for lexicography, have substantial power in managing language change. This is hardly the case because grammars and dictionaries have high authority but low reach (Zgusta 2006; Spolsky 2011; McLelland 2021: 283). Less acknowledged are competing (political) motivations and polarized (professional) interests in language standardization. Estonia is not an exception.

Institutionalized language planning in Estonia began with the work of the Society of Estonian Literati (1872–1894), though notable metalinguistic texts were published earlier. The first systematization efforts for Estonian included the comprehensive Estonian-German dictionary (1869) and descriptive grammar (1875), compiled by Ferdinand Johann Wiedemann, an Estonian-born linguist of German-Swedish origin. Wiedemann was not seeking to create a standard, but to bring together the different strands forming Estonian – to organize Estonian and make it visible (Wiedemann 2014: 391 as cited in Ross 2019). The republished editions of the dictionary released as late as 1973 testify to the magnitude of his work.

Between 1908 and 1911, two non-state associations – the Estonian Popular Education Society (1906) and Estonian Literary Society (1907) – convened four conferences to standardize written Estonian. This also signaled the start of the collective effort to compile the first Estonian dictionary of correct usage Eesti keele õigekirjutuse-sõnaraamat (Raun 1985). The dictionary was published in 1918, shortly after Estonia gained independence (Päll 2019). While not entirely prescriptivist, it excluded vernacular words and neologisms but retained parallel forms – its aim was both national and nationalist, which was typical for the long(er) nineteenth century (see Rutten 2009; McLelland 2021).
Between the world wars, but also post-WWII, standardization was entirely the non-state business of language professionals, incl. the working group of the Estonian Literary Society, the Mother Tongue Society (est. 1920), and the Soviet equivalent of the latter – the National Orthology Commission (est. 1960). While adhering to pre-war prescriptivism, the 1945 and 1946 reprints of the Concise Orthographic Dictionary (1933), for example, contained several hundred Soviet neologisms and omitted ideologically contested words used in the independent pre-war Republic of Estonia; however, its orthography, orthoepy, and morphology remained largely intact (Erelt 2002: 155).

The prescriptivist stance eased up towards the late 1970s and 1980s (see Päll 2019). The Estonian national language planning agency Eesti Keele Instituut (EKI; The Institute of the Estonian language), established in 1947, oversees the Estonian Literary Standard. This standard is the system of spelling, grammar, and lexical norms and recommendations detailed in the most recent Dictionary of Standard Estonian (DSE). The standard also includes the decisions of the linguistic committee of the non-governmental Mother Tongue Society as well as the orthographic rules, normative handbook, and grammar approved by its language committee (Language Act § 4).¹ As a result of the Language Act of 2011, the DSE embraces both corpus and status planning agendas (see Fishman 2006).² Beyond its language standardization goals, the EKI has other tasks, incl. linguistic and onomastic research. Since 2018, the EKI has been working towards the ambitious goal of merging a number of different dictionaries into a single “superdictionary” – the EKI Combined Dictionary (CombiDic) (e.g., Tavast et al. 2020).

Instead of linking dictionaries or setting up an aggregated search mechanism across dictionaries, the EKI is creating a user-friendly application by assuming that people look for words without scanning through different dictionaries. Also, as CombiDic will be usage- or corpus-based and descriptive, it will diverge from the prescriptive approach, which has dominated corpus planning until recently (Langemets et al. 2021:)

¹ The procedure for the application of the Literary Standard in written language shall be established by the regulation of the national government. “Eesti kirjakeele normi rakendamise kord”, Riigi Teataja I, 2011, 3.
² While Klöss (1969) made a clear distinction between status and corpus planning, both have also been seen as “just two sides of one and the same coin” (Fishman 2006: 17).
Therefore, lexicographers are actively seeking a solution for how CombiDic could describe language use but also provide standardizing recommendations (Langemets et al. 2021). The launch of the “super-dictionary” has provoked a public debate in (print) media. In it, different language professionals – mostly those who are responsible for the implementation of Standard Estonian – have expressed their concerns over this change and the relationship between CombiDic and the DSE. Others, including lexicographers and linguists, introduced their agendas or acted in response to implementers’ concerns. This debate brought forth numerous explicit opposing viewpoints in the discourse of language standardization. The opposition between specialists and non-specialists or between the views of different specialists has been discussed by several scholars researching standardization.

This paper addresses the argumentation strategies of different language professionals – linguists, lexicographers, proofreaders, copy editors, and educators – with a broad research question describing some of the types of extralinguistic arguments voiced in this debate. Standardization may have merely linguistic goals, but its underlying principles are often extralinguistic, mostly social. Those non-linguistic aims of standardization may involve purism, prestige, nationalism and national unity, aesthetics, economic advancement, and (de)colonization (for an outline see Kaplan & Baldauf 1997: 38; Deumert & Vandenbussche 2003: 5). The discursive aspects of Estonian corpus planning have been empirically explored by only a few scholars (e.g., Raag 1999; Laanekask 2004; Undusk 2012; Vainik & Paulsen 2023) who have not openly positioned their analyses within critical studies. Critical inquiry reveals how social differences, which stigmatize some speech forms or varieties and rank languages and their users according to their perceived value, are constructed in LPP. A discursive approach is needed to understand the workings of LPP in creating social inequalities because people behave and interact according to socially constructed meanings not grounded in evidence-based, scientific reasoning. This paper has two parts: theoretical and empirical. In the first part, the ideological and political foundations of standardization are reviewed and discursive approaches to LPP are discussed. In the second part, findings from discourse analysis are presented and interpreted, and conclusions are drawn.
2. Language standardization: 
a special type of sociolinguistic change

For Valter Tauli, the internationally renowned Estonian theorist of language planning, an ideal language was economical, i.e., “by the minimum of means [it] attains the maximum of results”, it is purposeful, elastic, and has an aesthetic form (Tauli 1974: 59–60). Some of those qualities, aside from being optimal (e.g., fewer, shorter words), involve judgements. For example, in Wittgensteinian philosophy, aesthetic judgments become normative claims requiring justification due to a particular alignment between grammatical and aesthetic systems (Appleqvist 2023: 28). Tauli’s theory of language planning is an applied and normative science with language planning viewed as an activity for regulating and improving a language or creating a new one.3

Nevertheless, his far-reaching instrumentalism failed to acknowledge that “any attempt of language planners to alter a language in the name of some ideal principle has to take into account the process whereby language is transmitted and maintained” (Haugen 1971: 273). In Haugen’s revised model of standardization (1983), language planning involves four stages: 1) selection of a norm, i.e., deciding on a particular variety; 2) codification of the norm with normative texts such as dictionaries and grammars; 3) implementation of function using political measures in order for prescriptions to be respected, and 4) elaboration of function by coining vocabulary and terminology for new uses. The first two stages include decision making (selection, status planning) and standardization procedures (codification, corpus planning). By producing normative (prescriptive) dictionaries, a normative grammar, and an orthography, standardization activities establish norms. Later, Haugen (1987: 63) also admitted that his model overlooks the non-linguistic goals of standardization.

As a human intervention into language change, standardization has long been understood as a specific type of sociolinguistic change (Milroy 2001; Deumert & Vandenbussche 2003: 1). It is the process of imposing a supradialectal norm onto a community’s varying linguistic practices with this norm commonly promoted by the state and taught in its, mostly centralized, educational system (Costa 2020: 1). Beyond

---

3 Emphasis added.
compulsory education, there are other techniques by which the state supports standardization such as censuses and map making (see Appadurai 2005). In time, this register comes to represent “the best and most legitimate form of a given language” (Costa 2020: 1; Milroy 2001). Being “first and foremost a sociopolitical phenomenon”, but also a linguistic process, which reduces variation (Deumert & Vandenbussche 2003: 2), standardization has attracted scholarly attention from both linguists and social scientists. Researchers within the tradition of North American linguistic anthropology first started to question how linguistic phenomena are invested with value and ranked by the (re)production and contesting of conventional indexical ties between features, genres, or varieties of language as well as representations of speakers in terms of ethnicity, gender, aesthetics, and morality (Milani & Johnson 2009: 4; Silverstein 1976, 1996; Woolard 1998; Irvine & Gal 2000). Soon sociolinguists joined the critical project of historicizing linguistic research by insisting that language involves complex systems of semiosis that exist in “sociolinguistically stratified economies of signs and meaning” (Blommaert 2013: 50). There is a growing body of scholars who address not only the issues of standardization and its legitimacy as well as the authority and authenticity of standards, but also how at a particular historical moment standardization becomes naturalized (e.g., Milroy 2001; Sebba 2007; Lane, Costa & de Korne 2017; Kristiansen 2019; Costa 2020; Ayres-Bennett & Bellamy 2021).

2.1. Three approaches to linguistic standardization

There are three broad approaches to linguistic standardization: technical, political or ideological and historical (Costa 2020). The first draws on the works of Einar Haugen (see Section 2 above), whose approach serves as a post hoc rationalization of how languages ended up being standardized but says little about how large speaker populations became convinced about the need and legitimacy of the standard (Costa 2020: 2; Haugen 1987). The second approach describes standardization as an ongoing ideologically-motivated project: a part of the standard language or monoglot ideologies of nation states, language genealogies,

---

4 The imposition of uniformity upon a class of variable objects is not a universal global phenomenon (Milroy 2001).
and social stratification (see Bourdieu 1991; Silverstein 1996; Milroy 2001). Standardization amounts to crafting what variety speakers (and learners) consider to be the language (Costa 2020: 2), i.e., the standard stands for *languaging*. The third approach raises the question of whether all linguistic norms, incl. spoken, count as the standard or whether the authority of the standard derives from a fixed – literary, legal, or regal – language such as a Bible translation (Costa 2020: 2; see Joseph 1987; Bauman & Briggs 2003). The appearance of new populaces within the realm of science and politics in the latter half of the eighteenth century and nineteenth century, respectively, occasioned “a quest for ways of speaking that were unambiguous, universal, and devoid of obvious indexicalities of place or religious affiliation in particular – in other words, *decontextualized* forms of language” (Costa 2020: 3). The current analysis of extralinguistic arguments in standardization is informed by the second approach, but many such analyses infuse both ideological and historical approaches.

### 2.2. Language standardization from above and below

In most language historiographies, the focus has been on printed, formal, and literary texts by elites; those texts make up the bulk of a standardized variety, which are top-down implemented by authorities (Rutter & Vosters 2021; Elspaß 2021). Other texts, mostly informal and handwritten, have been long ignored by (historical) (socio)linguists and language planners (for lower-class writing see Vandenbussche & Elspaß 2007; Elspaß 2021). Moreover, texts, which fall between spoken and written modes, e.g., language used in the internet communications or speech produced by non-elites or stigmatized groups such as youth or non-native speakers, have also been long ignored by linguists (see Labov 2001; Eckert 2003; Androutsopoulos 2008).

These two views of language planning exist in parallel with (1) structuralist and generativist schools in linguistics, which ignore language performance and consider language to be a system separate from human interaction and cognition, and (2) usage-based linguistics, which engages with language development, be it diachronic change or

---

5 *Languaging* refers to the Humboldtian idea of language as a process: a process of meaning-making and shaping knowledge via experience.
individual growth in acquisition (von Mengden & Coussé 2014; Diessel 2017). It is equally important that the usage-based approach emphasizes the role of frequency in language (more is more), incl. for semantic change (Bybee 2007; Gipper 2014). Language standardization from below is about the conventionalization and acceptance of language variants and varieties, in other words, it “rests on the assumption that potentially all members of a language community and all of their forms of verbal interaction have contributed to the standardization processes of this language” (Elspaß 2021: 93–94).

2.3. Language planning as discourse

Discourses are practices which are conditioned by the regularities of a particular discursive formation and “form the object of which they speak” (Foucault 1972: 49). In other words, discourse is a social practice of meaning-making (Martin & Rose 2003). The key ideological function of discourse is legitimation; legitimation is a form of collective action which seeks to justify itself (van Dijk 1998). Discursive practices may (re)produce unequal relationships between groups, by representing and positioning some phenomena, such as linguistic practices or forms, of greater worth than others. To gain a more nuanced understanding of LPP, standardization, and lexicography, critical and discourse-analytical approaches have proven to be more than appropriate (Blackledge 2005; Koreinik 2011; Paffey 2012; Barakos, Unger 2016; Hult 2017; Heller, Pietikäinen & Pujolar 2017; Wenge 2019). Critical Discourse Analysis (CDA) has been utilized by scholars interested in unmasking abuse of power and the ideologies and interests of particular groups (e.g., van Leeuwen 1995; van Dijk 1998; van Leeuwen & Wodak 1999). A deconstructive activity such as CDA can play a key role in reversing negative language ideologies adopted by some groups of speakers towards their forms of speech or writing and should be incorporated into positive language planning and language policy projects (Lo Bianco 2009: 101). For Lo Bianco (2008: 157), language planning is:

6 Emphasis added.
an ensemble of activities, some of which are textual (laws, reports, authorizations), others of which are discursive (speeches, radio debates), while still others involve the public performance of behaviors that powerful individuals or institutions hold up as models to be followed in the analysis which [--] is applied to three levels of policy activity [--], the intended, the enacted and the experienced.

Thus, the instances of language planning to strive for include: public texts such as constitutions and other declarative texts (both practical and symbolic), public discourses or “ongoing debates, discussions and arguments on issues of languages”, and performative action, i.e., the very practice of language use (Lo Bianco 2009: 102). That said, language planning bears a resemblance to “a motivated conversation” about how a speaker population “talks to itself about its principal communicative resource, i.e., language” because language change requires “mass validation in practices of language use” (Lo Bianco 2009: 102).

3. Research design: data and method

To explore what extralinguistic arguments are used in this language planning discourse, I have collected newspaper texts on the macro topics “superdictionary” and “dictionary reform” from Estonian newspapers with national distribution since 2021, then qualitatively analyzed these texts using CDA (for qualitative discourse analysis see Martin & Rose 2003). Without attempting to establish the typical, qualitative discourse analysis explores the meanings beyond the clauses as well as the relationship between discourses and of discourses with social situations.

Texts were selected with a simple online search using the search words: EKI ‘Institute of Estonian Language’, ÕS ‘Dictionary of Standard Estonian’, eesti kirjakeel ‘Estonian literary language’, keele­korraldus ‘language planning’. Most of the texts referenced or were linked to one or several other texts on the same topic.7 However, without tracing all existing texts on each topic, including unpublished texts or those published offline, it is impossible to state explicitly the size of the text population. Nevertheless, I ended up with a small text corpus.

---

7 The first text in my sample is from January 29, 2021, where language planners explain the reform (see MLPP 290121).
(see Data sources), incorporating one blog entry posted by Päevakera, which had been picked up and republished by the news portal of Estonian Public Broadcasting (EPB), seven opinion pieces first published in the weekly newspaper Sirp and then republished in EPB, and three opinions, commentaries, or columns published by EPB. Two articles on the topic were also included in the corpus from other, privately-owned print and web media outlets with considerable readership such as Postimees and the online portals Delfi and Õhtuleht (owned by Postimees Group and Ekspress Group, respectively), which restrict access to their content with a paywall. Texts vary in length and genre; interviews were excluded. Apart from receiving public attention and having political influence, print and online media have practical and other substantial advantages over broadcasts: there is no need for transcription or speech-to-text technologies, these media are disseminated to large audiences, their texts reflect the social mainstream (Mautner 2008). However, I am fully aware of the opportunities and challenges of using web-based data for CDA.8

To analyze extralinguistic arguments in this text corpus, I have employed the discourse-historical approach to critical discourse analysis (CDA), which has also proven to be a useful tool not only for the analysis of different legitimizing efforts (van Leeuwen & Wodak 1999) but also manipulation (van Dijk 2006). In general, legitimation strategies are mainly employed to defend and preserve problematic narratives or past controversial actions (De Cillia, Reisigl & Wodak 1999). My analysis primarily focuses on legitimation and argumentation strategies within language planning discourse. Van Leeuwen and Wodak (1999) outline four macrostrategies for argumentation: constructive strategies, strategies of perpetuation and justification, strategies of transformation, and destructive strategies.

This analysis is limited to perpetuation and transformation strategies, which either justify or seek to alter the status quo, respectively. Both of these strategies involve the “grammar of legitimation” (van Leeuwen 1995): authorization, rationalization, moral evaluation, and mythopoiesis. Authorization uses references to history, tradition, law, religion

---

8 Mautner (2005: 815–819) highlights the size of the web as both a blessing and curse, due to its diversity of voices, its dynamic and ephemeral quality, and issues related to textuality, interactivity, and multimodality.
(most often impersonal authorization), and people (personal, conformity authorization). Furthermore, (de)legitimation discourse semantically refers to the past actions of others for which the actor takes responsibility (Rojo & van Dijk 1997). Additionally, comparison – “the claim that (legitimate) others have engaged in similar actions” – is explored in both – perpetuation and transformation – strategies (Rojo & van Dijk 1997: 537).

Furthermore, referential (and predicative) strategies are used to represent allegiances, memberships, and groups. Opposition and inter-group polarization, without being my primary foci, are represented by collective nouns (people, youth, researchers, Estonians) or by reduced referential devices such as personal pronouns (us vs. them), which create social distance (van Dijk et al. 1997; Wodak 2003). The use of referential devices is not only conditioned by the context, but also creates one (Silverstein 1976). The referential and predicative function in a clause is performed by a noun phrase (NP). For example, the English language offers several NP types, each of which denotes the plurality of entities or semantic plurality (aggregate nouns, partly substantivized adjectives, conjoined NPs) (Gardelle 2019). In Estonian, the referential (and predicative) function of plurality is performed also by a quantifier phrase.

4. Findings and discussion: the grammar of legitimation

4.1. Authorization strategies

The first strategy is personalized authorization, which in all but one case also involves reference to history and tradition; references to academics or other learned authorities are also used (excerpts 1–3). There are plenty of additional instances voiced in other analyzed texts (e.g., HM 220222). References to learned authorities are used by both “pro-reformers” and their counterparts, which in turn refer to past oppositions in that matter.

---

9 There is an obscure line between referential and predicational strategies; the latter involves the attribution of traits, characteristics, qualities to social actors (Wodak 2003).
A historic change of direction for “ÕS 1999” initiated and led by Henn Saari [1924–1999; the Estonian linguist and language planner] has now, twenty years later, helped to pave the way for a changing ÕS (MLPP 290121).10

This situation [of over-regulation] causes defiance and sometimes unwillingness to participate in the Estonian-language space, especially when one has a choice, as, for example, Jaan Kaplinski had [1941–2021; one of the most well-known and translated Estonian writers] or as most of today’s youth, unfortunately, have due to their good English skills (LL 290422).11

A human is born in the middle of texts, acquires language from texts, and communicates through the medium of texts, as Juri Lotman emphasized [1922–1933; the semiotician and literary theorist] already 40 years ago (KK 210722).12

Another strategy is comparison with legitimate others (cf. role models); a claim of others being involved in similar practices is made (excerpt 4). Europe has been serving as a legitimate point of reference also in other (Estonian) language ideological debates (Koreinik 2011). In excerpt 5, the DSE is compared to institutions and/or symbols of national significance where knowledge is accumulated over centuries.

European ethnicities have formed largely based on language and Estonians are no exception (AT 110522).13

The DSE should be seen as an institution of national culture comparable to the National Library [of Estonia] or the University [of Tartu] (KT 040621).14

10 “Henn Saari algatatud ja juhitud ajalooline suunamuutus „ÕS 1999“ jaoks on näüd, kakskümmand aastat hiljem, aidanud omal moel taas teed sildutada muutuvale ÕSile”.
11 “See olukord tekitab aga trotsi ning vahel ka soovimatust eesti keeleruumis osaleda, eriti kui on valida, nagu näiteks oli Jaan Kaplinski või on nende hea inglise keele oskuse taustal paraku suuremal osal tänapäeva noordest.”
12 “Inimene sünnib tekstide keskele, omandab tekstidest keele ja suhtleb tekstide vahendusel, rõhutas Juri Lotman juba umbes 40 aastat tagasi.”
13 “Euroopa rahvused on kujunenud suuresti keele alusel ja eestlased ei ole mingi erand.”
14 “Kuigi selle kohta ei ole eralda seadust, peaks ÕSi vaatlema kui rahuuskultuurilist insti-tutsiooni, võrreldavat rahuusraamatukogu või ülikooliga.”
Alongside comparison with role models are metaphorical expressions which compare language with an organism (or its parts, such as the brain or bloodstream; cf. *pars pro toto*) (ML 180621), language loss with the loss of natural diversity, or corpus planning with clearcut logging (see excerpt 12 below) (see also Vainik & Paulsen 2023). The next comparison (6) also functions as a *predicational strategy* and *negative other-presentation*. It represents the proponents of the “super-dictionary” as tech-savvy people who do not see a future for paper/print books.

(6) The worship of the language corpus as the source of truth has led the progress-believing liberals to the conclusion that the language corpus is an internal combustion engine, and the dictionary is a draft horse whose breed must be swept off the earth by the tidal wave of the revolution (KT 040621).

4.2. Rationalization strategies

Rationalizations may include causal *schemas towards the past* or *future* and are presented as apparent facts. Rationalizations are often linked to moral values. *Instrumental and theoretical rationalizations* include the justification of practices by reference to their functions; activities, which are (de)legitimated are subject to cause-and-effect relationships (van Leeuwen & Wodak 1999). Excerpt 7 illustrates a *causal schema towards the past*, referring to the past activities of corpus planners, which have proven ineffective:

---


16 “lugu sarnaneb metsandusega”, “‘keelemetsas’ võib sõnastikureform tähistada lage-raiet”.

17 “Keelekorpuse jumaldamine tõe allikana on progressiusksed liberaalid viinud järel-duseni, et keelekorpus on sisepõlemismootoriga sõiduk ja sõnastik veohobune, kelle tõu revolutsiooni hiidlaine peabki maamunalt pühkima.”
Despite the fact that there have been efforts to make these [paronyms] clear to [language learners] for nearly a hundred years, today’s language users do not know how to use these in the way that the norm was created (LL 290422).

Causal schemas towards the future are also used. These refer to unwanted outcomes such as threats or danger and are used by both the proponents and opponents of the “dictionary reform” (see excerpt 2). For example, the latter voice sees danger in digital dictionaries being vulnerable, crashable, and hackable. In excerpt 17 below, the threat that literary Estonian could be destroyed by the usage-based lexicographic approach to which the new dictionary adheres is voiced by a columnist from a mainstream news outlet who is also identified as a historian. Another, but different, danger, which relates to speakers’ agency and ownership of their language, is voiced by linguists (excerpts 2 and 8). A view expressed by the director of the EKI, and echoed by a professor of modern Estonian, is that a real, serious, and existential danger exists in overdetermination, as it may lead young people to choose English for academic writing or other texts.

This myth [“that there is a centrally planned more accurate and appropriate way of speaking and writing than our mother tongue”] is dangerous to Estonian as it suppresses speakers’ initiative and ingenuity in the use and advancement of their language and deprives them of the responsibility that comes with it (AT 110522).

The causal schema towards an unwanted future, but also other examples from the text corpus belong to the discourses of language endangerment (see Koreinik 2011), which is not surprising given that

---

18 “Hoolimata sellest, et neid on ligi sada aastat üritatud õppijale selgeks teha, ei oska ka praegused keelekasutajad neid kasutada nii, nagu norm loodi.”
19 “Digisüsteem on haavatav, see võib blokeerida, kokku kukkuda, rääkimata häkkimisest, viirustest jms. [--] Eesti keele ajalukku tekib sügav auk.” (ML 180621)
20 “See on reaalne, tõsine ja eesti keele jaoks võib-olla eksistenttsiaalne oht – kui noortele tundub, et neil on lihtsam oma akadeemilisi või muid teksti kirjutada inglise keeles, sest „inglise keel on lihtsam, seal on vähem reegleid.“” (LL 290422)
21 “See müüt on eesti keelele ohtlik, kuna surub alla kõnelejate initsiatiivi ja leidlikkuse oma keele kasutamisel ja arendamisel ning jätab nad ilma ka sellega kaasnevast vastutusest.”
concerns regarding the endangerment of Estonian have been voiced since the mid-1800s (Ehala et al. 2014).

**Theoretical rationalizations** usually include explanations and definitions; the latter are used to define the language as such, the Estonian language, common Estonian, and literary Estonian (excerpts 9–11).

(9) A good human language should unite and differentiate, be common and personal, stay intact and adapt to changing circumstances, be grounded in agreement but work without agreement. [---] Language [languaging] is not about following rules while being afraid of a red pen. Language is a core element of being human, which facilitates the bulk of our lives and which we use to define ourselves. Language is a source of joy and playfulness (AT 110522).

(10) Language is a complex, essentially democratic, and element-like [στοῖχος (stoîkhos) + -eîon (-eîon)] system, which resists being compartmentalized into “right” and “wrong”. [---] Estonian is much more than [its] written form and commas (LL 2904022).

The latter definition of Estonian represents a structural metaphor – language as a tightly organized system of mutually-related parts (see Haugen 1972) – and questions the domination of written over spoken language. The issue of writing being privileged over speech is explored by Blommaert (2004: 644), who suggests revisiting the functions of written texts and recovering orality as a function of literacy. He also notes that “literacy is not just part of “language” in general; it is a specific manifestation of language use, related to spoken language but different as a field of action” (Blommaert 2004: 644). Graphocentrism and textualism are powerful language ideologies that organize a great deal of what is found in language and how language is believed to function.

---

22 In Literary Estonian, ‘literary’ is not synonymous with either the Estonian Literary Standard or Standard Estonian. In popular belief, it can mean both an ideal Estonian or common Estonian, depending on the viewpoint.

23 “Üks hea inimkeel peaks üheaegselt ühendama ja eristama, olema ühine ja isikupäranne, püsimu muutumatuna ja kohanema muutuvate oludega, põhinema kokkuleppel ja töötama ka kokku leppimata. [---] Keel ei ole reeglite järgimine punase pastaka hirmus. Keel on inimeseks olemise keskne osa, mille kaudu käib suur osa meie elust ja mille kaudu me ennast määratleme. Keel on rõõmu ja mängulusti allikas.”

24 “Keel on keerukas, olemuslikult demokraatlik ja stiihilise süsteem, mis ei allu “õigeks” ja “valeks” lahterdmisele. [---] Eesti keel on palju rohkemat kui kirjapilt ja komad.”
(Collins 1996). Excerpt 11 captures the belief well that the standard and texts written in the standard stand for the language and therefore it is paradigmatic of those ideologies (see Costa 2020).

(11) Language does not exist in any other form than as a text – we also abstract Estonian from the texts of our language community (KK 210722).

Another fascinating distinction is made between Common Estonian and Literary Estonian. The former is a shared part of the linguistic experiences of all members of the Estonian language (or speech?) community. The latter is related to the DSE which establishes the Estonian Literary Standard: on one hand, it is the legal ground for the literary norm, but on the other, it is professionally compiled and has a long tradition. The negation can also function as a definition “opinions or comments in social media are not the literary language yet, but a totally new language variety.” These and the next theoretical rationalization by a professional copy editor seem to support an understanding that linguistic exchanges, incl. spoken ones, listed in this quotation, occur in the Estonian Literary Standard.

(12) I remind [you] that the standardized literary language is taught at Estonian schools and universities, [Estonian] public broadcasting is working, and laws are written, our state apparatus functions [in it], etc. (MK 300322).

25 “Keelt ei ole olemas muul kujul kui tekstina – abstraheerime ka eesti keele oma keelkogukonna tekstidest”.
26 “Ainult kõigi ühe keelkogukonna elavate liikmete keelekogemuse ühine osa on eesti ühiskeel” (KK 210722).
27 The Literary Standard is determined by the most recent dictionary of correct language use published by EKI, decisions made by the language committee of the Estonian Mother Tongue Society as well as the orthographic rules, normative handbook, and grammar approved by it (Language Act § 2).
28 “ÕS on ühest küljest kirjakeele normi seaduslik alus ja teisest küljest kirjakeele suure traditsiooniga professionaalselt koostatud alus” (KK 210722).
29 “[−], sest arvamused või kommentaarid ühismeedias ei ole ju kirjakeel, vaid täiesti uus keelevorm” (ML 180621).
30 “Tuletan meelde, et normitud kirjakeele õpetatakse Eesti koolides ja ülikoolides, töötab rahvusringhääling, kirjutatakse seadusi, toimib meie riigiaparaat jne” (MK 300322).
4.3. Moralization and mythopoesis

Legitimation can be also realized by mythopoesis or storytelling. In a moral tale, a protagonist is rewarded either for engaging in legitimate social practices or restoring the legitimate order; in a cautionary tale, a protagonist engages in activities which lead to an unwanted ending. There are two voices, which explicitly represent one of the two tales in their texts. Beyond other strategies, the first author tells a self-narrative about a social gathering, where the protagonist meets someone with a doctorate who recently discovered that Estonian does not have strict rules, i.e., the impression (s)he got from her/his proofs being “repeatedly proof-edited in red” (LR 070922).

There are several comparisons represented in this text, e.g., language is compared to music (“Does the user of a language have to bend to the rules, or does a language adapt to its users? This would be equivalent to asking whether someone who listens to music should only be allowed to listen to the tunes played on a particular radio station.”)\(^\text{31}\) given that the author positions herself as a linguist and an occasional DJ. Another comparison is first found in the title, where language is compared to a low-cost apartment building built in the Soviet Union during the early 1960s. The main argument is a moral one and compares the fixation on the standard with living forever in an old and dysfunctional building, which must not be renovated, and its space kept unchanged, its walls unpainted and solar panels not placed on its roof.\(^\text{32}\)

Another story, a cautionary tale, starts with the opening “Welcome to the garden I take care of!” and compares the (Estonian?) language with a biologically diverse but well-maintained garden, which, however, needs to be protected against weeds, alien species, and bird damage.\(^\text{33}\)

The protagonist is a gardener, whose tasks include watering, weeding, supporting some plants but also thinning and rooting out others, and

---

31 “Aga kumbapidi siis ikkagi on: kas keelekasutaja peab painduma reeglite järgi või kohaneb keeloma kasutaja järgi? Umbes samaväärne oleks kūsida, kas muusikakuulaja peab kuuluma ainult lugusid, mida mängib mingi teatud raadiojaam.”


33 The same author tells another anecdotal story about an Irish prostitute to demonstrate why it is important to plan semantics (see HM 220222).
eliminating alien species. Yet, the protagonist does not own the garden, which is used daily for substantial and aesthetic ends (cf. instrumental rationalization, goal orientation) (van Leeuwen 2008: 113–114). The owner of the garden has left the protagonist without the necessary tools and has hinted that biodiversity is preferred from now on. The narrative ends with remembering the many years of gardening which gave the protagonist work, pleasure, beauty, and cooperation, but which sadly must end as biodiversity does not require care. The same metaphor is used by another author, the editor-in-chief of a cultural periodical, which aims to act as an intermediary between the most recent standard and development of different academic fields. The term “unwanted plant” is used for a redundant linguistic form or language variety: “if a language remains unprotected, the weeds will take over” or “the EKI is not planning to protect language anymore, but allows the weeds to run wild” (TK 141122). English loanwords and other contact phenomena are defined as unwanted plants:

(13) If from now on those thousands of flowers or plants grow from seeds flown here from somewhere in America. It would mean that instead of a cultivated park or garden plots, a thicket would be permitted where alien species (- loans) run riot. [--] It is not certain whether this language-thicket springing up thanks to the winds of the world will take root in Estonian at all (TK 141122).
Maffi (2005) reviewed interdisciplinary research into the world’s linguistic, cultural, and biological diversity as manifestations of the diversity of life. Some of the authors compare species with languages and others indicate a remarkable overlap between linguistic and biological diversity throughout the world. Her examination starts from the Darwinian parallel between evolutionary biology and historical linguistics. From the 19th century onward, historical-comparative linguistics has been the school of language sciences which has used the biological metaphor – language as an organism that lives, reproduces, and dies (Haugen 1972: 326). The cautionary tale above and referential strategies described below go further by equating language with a garden/forest and representing (cf. referential strategy) corpus planners as gardeners or harvesters (see excerpt 12) (see also Vainik & Paulsen 2023). Another, instrumental metaphor – language as an instrument or a tool – is employed in the title of text “Let Estonian be a comfortable tool” (LL 290422), which is an example of theoretical rationalization (comparison).

5. Findings and discussion: Referential strategies

There are plenty of ways to represent social actors: exclusion, role allocation, classification, categorization, nomination, indetermination (van Leeuwen 2008). Mostly representational choices personalize social actors by representing them as human beings; those are realized by proper names, nouns or personal or possessive pronouns (van Leeuwen 2008: 46). Pronouns are political words, which discursively create intergroup polarization (cf. “us” vs. “them”) (Pennycook 1994); political discourse is also the discourse of identity and difference (Wodak 2003).

More or less neutral referential devices are used in this discourse depending on whether it is voiced by opponents or proponents of “the reform”: “the liberators of literary language or the initiators of the dictionary reform”, “progress-believing liberals”, “academic language informants”, “the liberal wing in language planning” (all KT 040621).

---

38 The evolutionary parallel backed by the belief in progress is often assigned to the Fergussonian (1968) three-component – graphization, standardization, modernization – corpus-planning process.

39 “Eesti keel olgu mugav tööriist.”
“the owner [of the garden]” (HM 160921), “reformers” (AK 040621), “the manager of the entire reform” (LV 241122), “the innovators” (KT 040621), and “language gurus” (LV 241022). Sometimes, those people are referred to with proper names and their professional roles (the EKI director Arvi Tavast, the EKI researcher of modern Estonian Margit Langemets, Prof. Liina Lindström). Opposing voices are (self-)referenced as “holders of occupations such as language teacher or language editor” (KT 040621), “I am not an opponent of corpus linguistics” (ML 180621). They also employ negative other-presentation by suggesting what the proponents might say: “philologists and other weirdos” (KT 040621). Excerpts 14 and 15 also include negative other-presentation: the EKI director is made responsible for abandoning the sustainable (forestry of) language planning and blamed for contradicting his own words.

(14) The question of the future of Estonian is no less sensitive than that of the forest and developers-innovators have much to do to 100% erase the doubts that “dictionary reform” may refer to clearcut logging in the language forest and that the EKI director is actually a harvesting machine from a production of the NO-theater [the state-owned repertoire theater NO99, which was active from 2005 to 2019 and the productions of which were known for their social relevance] (KT 040621).40

(15) Tavast, the manager of the entire reform, can say in one breath that there is no right or wrong language, only that [language] is spoken as it is spoken; however, saying norm instead of norming is still wrong. So, it only keeps getting clearer (LV 241022).41

(16) With that we approach the main trouble of the ongoing “creeping language reform”: it is inconsistent, illogical, and thus directly harmful to the literary language. [-] And Sõnaveeb ['Word Web'] is what is relied on by some especially snobby language gurus, who think it has to do

40 “Eesti keele tuleviku küsimus ei ole metsa omast sugugi vähemtundlik ja arendajatel-uendajatel on väga palju teha kustutamaks sajaprotsendiliselt kahtluse, et keelemetalas võib „sõnastikureform“ tähistada lagerait ja et EKI direktor on tegelikult harvester NO-teatri etendusest.”

41 “Kogu reformi juht Tavast oskab ühe hingetõmbega öelda, et pole õiget ega valet keelt, vaid kõneldakse nii nagu kõneldakse, aga normeeringu asemel norm õelda on ikkagi vale. Nii et aina selgemaks läheb.”
with something fabulously novel, youthful, and progressive, especially when compared to the out-of-date DSE, which is, alas, a whole four years old (LV 241022).  

The use of inclusive personal pronouns ("we”, “us”, “ours”) is variable as can be seen in excerpts 17 and 18 below. “We” and “our” are mostly used to refer to speakers of Estonian or the Estonian people: “in front of all our eyes” (MLPP 290121). Inclusive pronouns are less used by opponents, but plenty of nouns, both singular and plural, collective and not, are used by both proponents and opponents: (Estonian) youth/young people, grown-ups, individuals, researchers, university professors, (language) user(s), EKI management, EKI employees, people (and their elected representatives), language professionals and hobbyists, language-sensitive people, teachers, the parliament, scientists, the executors of state power, many people, linguists, a language learner, an (average) language user, a reader/listener, new generations, colleagues, the compilers of dictionaries, the Estonian-speaking population, sociologists, all language maintainers, language specialists. Excerpts 18 and 19 explicitly state who is “us”, i.e., not bloggers or criminals, not all who “open their mouths”, which in this context could also mean all speakers in Estonia. This is characteristic of language standardization from above, which ignores non-elite writing and informal language as inputs for a standard variety. The question remains who are those who proceed towards destruction. The use of a plural first-person pronoun in an exclusive manner creates social distance between those whose language is good enough for language planning and the rest of the speaker community instead of enhancing the community. While stigmatizing some people and their language use, those quotations remind readers that one of the goals of standardization has been to erase the indexicalities of social origins (Costa 2020).

42 “Millega me jõuamegi käimasoleva „hiiliva keelereformi“ põhihädani: see on järjekindlusetu, ebaloogiline ja seetõttu kirjakeelele otseselt kahjulik. [--] Ning just Sõna-veebile toetuvad mõned iseäranis ennastääis keeleorganid, kelle arvates on tegemist millegi vahvalt uue, noortepärase ja progressiivsega, iseäranis vananenud ÕSi kõrval, mis on, oh häda, tervelt neli aastat vana.”

43 “Meie kõigi silma all.”
If language folks are engaged with language every day, then outside our narrow circle its popularity comes in waves. Language is something natural for all of us, which simply works without special effort or even realizing it. It is important for us to speak Estonian.

Which brings us to the main problem of the ongoing ‘creeping language reform’: it is the incoherent, illogical, and therefore directly destructive to the literary language. We are broadly moving towards the target of the Estonian-language norm being shaped by everyone who can open their mouths.

This has caused disappointment: the EKI is not planning on protecting and purifying our (literary) language, i.e., to do language planning anymore; but it deals mainly with registering different linguistic phenomena, i.e., with analyzing language corpora, which brings about the equal treatment of all kinds of language use, regardless from where a clause got into the corpus: blogs, tweets, etc. There is a lack of classic literature in the language corpora which are starting to shape our language. This is how the language of bloggers and professional criminals will dominate instead of Alver’s [Betti Alver, 1906–1989, Estonian poet, prose writer, and translator] and Alliksaar’s [Artur Alliksaar, 1923–1966; Estonian poet, playwright, and translator].

Beyond representing social actors, it is equally important to understand how social action itself is represented. Often an action – lexicography and standardization activities informed by usage-based linguistics – is objectivated as “reform” or “the tidal wave of the
revolution” or with objectivated naturalization in the title “language casts off chains” (see also excerpt 13). All such expressions deactivate or deagentialize social actions (van Leeuwen 2008: 73).

6. Conclusions

CDA is a deconstructive activity and its importance lies in reversing destructive language ideologies adopted by some groups of speakers towards their forms of speech or writing (Lo Bianco 2009). By employing CDA as a research tool, extralinguistic arguments in the discourse of language standardization, which departed from the launch of a metalinguistic CombiDic “superdictionary” and the publication of the new edition of the DSE, were analyzed. Those arguments can also be regarded as language ideologies, i.e., “sets of beliefs about language articulated by users as a rationalization or justification of perceived language structure and use” (Silverstein 1979: 193). Besides surfacing the non-linguistic goals of standardization, indexes between the features, genres, styles, or varieties of Estonian and representations of its speakers are also (re)produced in this discourse (see Milani & Johnson 2010). Above all, the analysis explains how language planning and standardization are enacted and social inequality is created. To analyze argumentation in the corpus of eleven (news) media texts in public and private media outlets, I have used the method known as the grammar of legitimation (van Leeuwen 1995), but also searched for polarizing referential strategies such as the use of nouns and pronouns.

Both the proponents and opponents of “the reform”, with their professional roles as linguists and/or lexicographers and language correctors-editors or philologists, respectively, employ several authorization and rationalization strategies. These include comparison, references to unwanted outcomes, which, paradoxically, were related to losing language change or language shift depending on who made the arguments. Beyond language concerns, there is also a division of labor at stake: language professionals make implicit claims regarding jurisdiction, i.e., “claims to classify a [language] problem, to reason about it, and to take action on it” (Abbott 1988: 40). As distinctive to “standardization

47 “Keel vabaneb ahelaist. [--] Revolutsiooni hiidlaine.” (KT 040621).
from above”, not all speakers’ language is considered worthy of being collected for corpora and/or recorded in dictionaries by the opponents of “the reform”. Their arguments are truly extralinguistic, i.e., have nothing to do with language as such, but values attached to certain ways of speaking and writing.

The question – whose language is codified – raises a further question for whom language planning is done. The answer can be inferred from referential strategies such as the use of the plural first- or third-person pronouns, which can either build memberships or create social distance. When the proponents of the reform use first-person pronouns either for their organization or inclusively for the entire speaker community, the opponents use “our” or “our language” for Literary Estonian exclusively. The latter representations also correspond to “standardization from above”, which suggests the implementation of a standard variety selected from elite texts and text genres. This is in stark contrast with the proponents’ stance of “standardization from below”, which assumes that all members of a language community and all of their verbal forms of interaction add to standardization (Elspaß 2021: 93–95). “Standardization from below” shifts the focus of language planning to the primary forms of human interaction, which are mostly both oral and informal, “not edited according to prescriptive norms” and when written, then by common people (Elspaß 2021: 101). This is the central concern of opponents of “the dictionary reform” and the macrotopic in the discourse centered around the launch of “the superdictionary”.

Beyond extralinguistic arguments, the question arises regarding what language – a complex and central phenomenon in standardization – is. Language is directly or indirectly represented by metaphors (for an analysis of metaphors in this debate see also Vainik & Paulsen 2023), which either are known from biblical narratives (e.g. Paradise Lost, Golden Age) (cf. Labov 2001) or associated with different schools of linguistics: historical-comparative (language-as-organism), structuralism and generative grammar (language-as-tool, language-as-structure) (see Haugen 1972). One of metaphors used by the opponents

48 When Viht & Habicht (2022) describe the activities of the Society of Estonian Literati as an example of standardization from below, they miss an important difference between including all Estonian-language verbal interaction in language standards and undoing colonial linguistics driven by Baltic Germans or other non-native speakers of Estonian.
of “reform” is language-as-garden, i.e., language is imagined as a garden which requires cultivation (cf. language cultivation). One of the proponents uses a language-as-tool metaphor. Language is also represented as language-as-texts, which consists of writing the best form of the language (Blommaert 2004).

Discourse offers many prospects for future research: negation and the representation of agency deserve special attention. Furthermore, corpus linguistic methods could be used for a significantly larger text corpus also extended with respect to genre. I conclude by paraphrasing the late Gunther Kress (2018): in the twentieth century, all efforts were put into establishing a norm rather than embracing that which exists. The above discourse analysis explicitly demonstrates language ideological motivation and political polarization over similar foci in language standardization.

Acknowledgments

I am grateful to Dr. Külli Habicht, Prof. Kalevi Kull, Dr. Helen Plado, Dr. Kerttu Rozencevalde, and the three anonymous reviewers for their help and valuable comments. I owe a lot to Dr. Uldis Balodis for his diligent proofreading and English-language editing. This work is supported Estonian Research Council Grant PRG1719.

Data sources

142  Kadri Korenik


References


49 An English version of the text is used for quotations.


Märksõnad: keele standardimine, argumentatsioon, keeleideoloogiad, diskursus, keelekorraldus, eesti keel