

Sergejs Kruks

Dr. sc. inf., media and communication scholar; Rīga Stradiņš University

Dr. sc. inf., mediju un komunikācijas pētnieks; Rīgas Stradiņa Universitāte

E-pasts / e-mail: sergejs.kruks@rsu.lv

DOI: 10.35539/LTNC.2024.0056.07

Ambiguous Grammatical Forms in Latvian Corpora

Neviennozīmīgas gramatiskās formas latviešu valodas korpusos

Keywords:

ambiguous grammatical form,
corpus analysis,
critical discourse analysis,
Latvian language

Atslēgvārdi:

neviennozīmīga gramatiskā forma,
korpusa analīze,
kritiskā diskursa analīze,
latviešu valoda

Summary

Ambiguous grammatical forms (AGF) combine the characteristics of different parts of speech, and it is up to the receiver to decide whether a word is, for example, a subject, object, or predicate. When the sender does not provide explicit cotextual and contextual cues for interpretation, she assumes no responsibility for the content of the message but reserves the right to reject the receiver's interpretation as inadequate. In this capacity, AGF can be used to maintain hierarchical social relations between communicators. Hence, language is not a neutral means of communication, critical discourse analysis maintains. This paper seeks statistical support for this claim by testing the null hypothesis: If AGF are neutral carriers of meaning, the difference in their relative frequencies in the specialized corpora is not statistically significant. Analysis of the Latvian language corpora rejected the null hypothesis. The highest incidence of AGF was found in messages created by powerful social institutions: laws, doctoral dissertations, and news from the Internet media. Literary works, magazines, and everyday conversations contain the least number of ambiguous forms.

Kopsavilkums

Neviennozīmīgas gramatiskās formas apvieno dažādu vārdšķiru īpašības, un saņēmējam pašam jāizlemj, vai vārds teikumā ir, piemēram, subjekts, objekts vai predikāts. Neiekļaujot kotekstā un kontekstā skaidras atslēgas interpretācijai, sūtītājs neuzņemas atbildību par vēstījuma saturu, taču saglabā tiesības noraidīt sūtītāja sniegto interpretāciju kā neatbilstīgu. Šādu neviennozīmīgu gramatisko formu īpašību komunikatori var izmantot hierarhisko sociālo attiecību uzturēšanai. Kritiskā diskursa analīze pauž, ka valoda tāpēc nav neitrāls saziņas līdzeklis. Raksts meklē statistisko pamatojumu šim apgalvojumam, testējot nulles hipotēzi: ja neviennozīmīgas gramatiskās formas ir neitrāli saziņas līdzekļi, tad to relatīva biežuma atšķirība specializētajos valodas korpusos nav statistiski nozīmīga. Latviešu valodas korpusu analīze noraidīja nulles hipotēzi. Visbiežāk neviennozīmīgas gramatiskās formas parādās vēstījumos, kurus radījušas ar varu apveltītas sociālās institūcijas: likumos, doktora disertācijās un tiešsaistes mediju ziņās. Daiļliteratūrā, žurnālos un ikdienas sarunās ir vismazāk šādu gramatisko formu.

Introduction

Ambiguity in language arises because some grammatical forms combine the characteristics of different parts of speech. Mostly, the problem refers to derivatives from verbs. In new forms, verbs lose grammatical categories of part of speech (person, number, tense, mood, and modality) and acquire new ones. Nominalization, passive voice, indeclinable participle, declinable present passive participle, and subjunctive mood are examples of ambiguous grammatical forms (AGF) in the Latvian language. They can refer to a process or result, an antecedent or simultaneous action, a wish or future state, a reified abstraction, or an action. The receiver must consider the cotext and context before deciding whether a word should be understood as a noun, verb, adjective, or adverb. When interpretation cues are absent, the receiver may rely on her personal knowledge of the world. In strategic communication aimed at influencing others, AGF is an advantage for the sender who can reject the receiver's interpretation as contradicting the surface grammar or irrelevant in the presupposed context. In particular, the issue is acute in situations of unequal distribution of power. The authors of laws and policy documents have the power to prescribe behavior and impose sanctions. Using AGF in orders, the sender does not assume explicit responsibility for the content, but still the order requests the receiver's compliance. The receiver finds herself in a double bind situation with incompatible constraints: She must respect the order, but its exact meaning is impossible to explicate.

Language is not neutral, critical linguistics and critical discourse studies contend (Fowler, Kress 1979; Fowler 1991; Fairclough 2013; Titscher et al. 2000). Agents select from a variety of grammatical forms to achieve their specific goals in specific social interactions; consequently, the concrete language use reflects the distribution of power in social structures. This paper seeks a statistical foundation for such a claim in the specialized Latvian language corpora. The specialized corpora represent language use in various registers or various social interactions that assume different relations between senders and receivers: monologue and dialogue, requesting and begging, convincing and sharing opinion, commanding and suggesting. If AGF are neutral to social interactions, then the difference of their relative frequency in specialized corpora should not be statistically significant. In other words, oral conversations, private diaries, literature, news media, science, and law should demonstrate a similar level of ambiguity.

Ambiguous grammatical forms in Latvian

Ambiguous gram-

matical forms of the Latvian language were discussed in recent research on legal discourse (Kruks 2023). This section reiterates the discussion and describes some more ambiguity resources established during the current research.

Linguists distinguish between lexical (a word has more than one meaning) and structural (a phrase can be segmented in more than one way) ambiguity (Gillon 1990; Kennedy 2011). Derivatives from verbs are examples of lexical ambiguity, but first I will introduce a grammatical form that combines both kinds of ambiguity: noun in the genitive case. In Latvian, genitive is a very productive case declension: nouns in the genitive case can name subject, object, possession, attribution, quality, feature, substance, application, purpose, and can also be appositive and partitive (Kalnača, Lokmane 2021: 106–109). For example, (1) can be interpreted as “Paul has educated someone” and “Someone has educated Paul”.

(1) Lūk, tas ir Paula (GEN) audzināšanas (NMZ, GEN) rezultāts (NOM).

Well, this is the result (NOM) of education (NMZ, GEN) of/by Paul (GEN).

In the genitive case, the proper noun Paul can be either the subject or object of the proposition. To disambiguate the propositions, the receiver relies on the background information provided by cotext and context. Chaining nouns in the genitive case introduces structural ambiguity that entangles the interpretation even further. To distinguish between object, subject, and their characteristics, the receiver must segment the proposition applying personal knowledge of the world. In the balanced Latvian Language Corpus 2022 the longest cascade contains 16 nouns in the genitive case in a row:

(2) Pārcelt (INF) I. Liepu (ACC) no (PREP) Finanšu (GEN) ministrijas (GEN) Eiropas (GEN) Savienības (GEN) fondu (GEN) stratēģijas (GEN) departamenta (GEN) direktora (GEN) vietnieka (GEN) amata (GEN) Satiksmes (GEN) ministrijas (GEN) Investīciju (GEN) departamenta (GEN) direktora (GEN) vietnieka (GEN) amatā (LOC).

To transfer (INF) I. Liepa (ACC) from (PREP) the position (GEN) of Deputy (GEN) Director (GEN) of the European (GEN) Union (GEN) Fund (GEN) Strategy (GEN) Department (GEN) of the Ministry (GEN) of Finance (GEN) to the position (LOC) of Deputy (GEN) Director (GEN) of the Investment (GEN) Department (GEN) of the Ministry (GEN) of Transport (GEN).

In laws and policy documents, long cascades denote the positions of officials in public governance institutions and the smallest units of the latter. Even if (2) is not ambiguous, the reader should make an effort to structure the sentence into segments that denote meaningful entities. Proposition (3) is a current notion in media and political discourses, although the knowledge of the political context does not suffice to segment it unambiguously.

(3) Latvijas (GEN) kultūras (GEN) identitāte (NOM).

Latvia (GEN) culture (GEN) identity (NOM).

The proposition has two meaningful segmentations that result in different interpretations: “identity of Latvia’s culture” and “cultural identity of Latvia”.

Now we turn to the grammatical forms derived from verbs. Nominalization (NMZ) as a substantiated verb conceals the grammatical categories of the verb (person, number, tense, and mood), thus reifying dynamic processes. Concealing the agent, passive voice (PV) presents the outcome of the action as an effect of objective forces beyond human control. In Latvian, indefinite passive tense forms are derived using the auxiliary verbs *tikt* and *tapt* in a finite form and past passive participle; the perfect passive tense forms use the auxiliary verb *būt* in a finite form and past passive participle (Kalnača, Lokmane 2021: 260). Passive voice forms express generalized, regular or habitual actions, or impersonal statements of a fact (Kalnača, Lokmane 2021: 263). The variety of meanings allows the sender to present her subjective opinion as a generalization or a fact. These two grammatical forms, NMZ and PV, are usually discussed in the English-language literature on critical discourse studies, but Latvian has some other ambiguous forms.

The indeclinable participle (IDP) presents an action as causally related to the main action of the clause (Kalnača, Lokmane 2021: 298). Combining the characteristics of an adverb and verb, it can express mode, time, cause, purpose, circumstances of, or conditions for an action (Nītiņa, Grigorjevs 2013: 585). As the grammatical categories of the verb are lost, it is not always possible to distinguish between the result and the process, the antecedent and the simultaneity. The differentiation of meaning depends on the perfective or imperfective aspect of the original verb; however, not all verbs distinguish between these aspects. IDP allows the sender to imply causality, eschewing the specification of relations between two actions so that the receiver cannot tell between the result and the process. Proposition (4) is a goal that the government set in a policy document, the National Development Plan for 2021–2027.

(4) Efektīva (GEN) pierobežas (GEN) ekonomiskā (GEN) potenciāla (GEN) izmantošana (NMZ, NOM), stiprinot (IDP) sadarbības (NMZ, GEN) saišu (GEN) veidošanu (NMZ, ACC).

Effective (GEN) border (GEN) economic (GEN) potential (GEN) use (NMZ, NOM), strengthening (IDP) cooperation (NMZ, GEN) ties (GEN) formation (NMZ, ACC).

The IDP ‘strengthening’ can name a precondition (first strengthen the ‘cooperation ties’ and then ‘use effectively’); or a parallel process (someone ‘uses effectively’ while someone ‘creates ties’). Nominalizations and cascades of genitive turn the

goal 'to cooperate with neighboring countries' into a process 'strengthening the formation' that does not envisage a tangible result.

In absolute dative constructions, IDP names the action, and a noun or a pronoun in the dative case names the agent. Proposition (5) demonstrates that the absolute dative turns the agent and action in circumstances of another action performed by an anonymous agent.

- (5) Policijai (DAT) veicot (IDP) izmeklēšanu (NMZ, ACC), meitenes (GEN) jakas (GEN) kabatā (LOC) tika atrasta (PV) arī (PART) pirmsnāves (GEN) vēstulīte (NOM).

As/when the police (DAT) was conducting (IDP) an investigation (NMZ, ACC), a suicide (GEN) note (NOM) also (PART) was found (PV) in the girl's (GEN) jacket (GEN) pocket (NOM).

Subjunctive mood (SUB) is called 'optative mood' (*vēlējuma izteiksme*) in Latvian. It is an umbrella designation for various forms of expression of the relation between the content of the proposition and reality: subjunctive, conditional, and optative mood. Since these forms of verb share a single morphological marker in Latvian, it is difficult to tell the non-epistemic modality (projection of actions in the future) from the epistemic modality (sender's evaluation of the truth or attitude towards the proposition). This is a task that the government set for itself in the National Development Plan for 2021–2027:

- (6) Lai mazinātu (SUB) profilaktiski un medicīniski novēršamo nāves gadījumu skaitu,

In order to reduce (SUB) the number of preclinically and medically preventable deaths,

- (7) ir būtiski uzlabojama (DPPP) ne tikai veselības aprūpes pieejamība, bet arī kvalitāte.

is significantly improvable (DPPP) not only the availability of health care, but also the quality.

The reason for the existence of policy documents is that the government assumes responsibility to achieve the stated goal by the defined year. However, the text (6) is ambiguous. SUB can express both a goal that the government will attain in the future (non-epistemic modality) and merely a wish (epistemic modality). Furthermore, the prerequisite for the achievement of the target (7) is expressed by a declinable passive present participle (DPPP) instead of the debitive mood. DPPP can express possibility or necessity; linguists also recognize it as synonymous with debitive mood with the proviso that it has a 'semantically nuanced' meaning (Nitiņa, Grigorievs 2013: 577–578). Explicitly, proposition (7) does not state that the government will have attained the formulated goal.

The semi-declinable participle (SDP) is a control variable in the research. This active voice participle refers to actions occurring simultaneously with the main action of a clause. Retaining some grammatical categories of verbs (number and gender), the SDP has much less potential for ambiguity. In propositions with a shared agent, the IDP and the SDP are synonymous (Kalnača, Lokmane 2021). Rarely used in speech and writing, it can serve as an indicator of cultivated language. In the balanced Latvian language corpus 2022, IDP is used 18.5 times more often than SDP, whereas in the female prose corpus the rate is 0.72. Arguably, female authors make use of SDP to mark the gender of protagonists.

Hypothesis and Method

The statistical null hypothesis of the research: The difference in relative frequency of ambiguous grammatical forms in the specialized Latvian language corpora is not statistically significant. Five AGF are the variables chosen for the research: cascade of nouns in the genitive case, nominalization, passive voice, indeclinable participle, and subjunctive mood. The declinable passive present participle mentioned in the previous section was considered unsuitable for the quantitative analysis because it has a variety of non-ambiguous meanings, and its ambiguous synonymy with the debitive mood can be established in qualitative analysis. The semi-declinable participle is a control variable: supposedly it must be used less frequently in corpora rich in ambiguous grammatical forms.

The sample includes 19 corpora available online at korpuss.lv/ and korpuss.lnb.lv/ (in Table 1 corpora are grouped into clusters identified by statistical analysis). When corpus design permitted, specific registers were singled out as subcorpora. Subcorpora of poetry, drama, and political essays by the writer Rainis (1865–1929) were sampled from the corpus of his works; subcorpora of the national and regional press were separated from the balanced Latvian language corpus 2022; corpus of television subtitles was divided by genre (fiction and documentary); corpus of PhD theses was split by topic (natural sciences, social sciences, and philology; philosophy, art, and musicology were assigned to a single subcorpus). Large corpora of internet news media, parliamentary debate, and law were split into smaller chronological subcorpora. Altogether, the sample included 39 units.

The absolute frequencies of the variables were extracted in the NoSketch Engine program using the Corpus Query Language (CQL) codes. Extraction of four grammatical forms is straightforward since they are tagged in the corpora:

IDP: [tag="v..pu.*"]

SUB: [tag="v..c.*"]

PV: [tag="va.*"]

SDP: [tag="v..pp.*"]

Nominalizations are not identified by a special tag. NMZ denoting the processes that are the focus of this research, have suffixes *-šan-* and *-īb-* (Kalnača, Lokmane 2021: 131–137). However, two letters of the latter are insufficient to distinguish NMZ from other words. Extraction by the suffix *-šan-* alone returns satisfactory results to establish the trend (Kruks 2023), and it will be used here:

NMZ *-šan-* : [lemma="*.šāna"]

Regarding the cascades, tentatively, the minimal sequence has been set to three nouns, hence the short designation G3+. To avoid splitting cascades in portions, the full sequence is delimited by words other than nouns in the genitive case or by punctuation marks:

G3+ : [tag!="n...g."][tag="n...g.">{3,}[tag!="n...g."]

The extracted data were normalized per million words. By default, NoSketch Engine normalizes the data to the size of the entire analyzed corpus rather than to the selected subcorpus. The total number of words in each corpus and subcorpus was recalculated, excluding non-words from the count, and setting the minimal frequency to one occurrence. The normalized data were processed in SPSS. The two-step cluster analysis and descriptive statistics methods were used to summarize and describe the main characteristics of the subcorpora and corpora and to identify patterns and relationships between them. Pearson's bivariate correlation establishes the co-occurrence of grammatical forms in corpora.

Findings

The two-step cluster analysis groups the corpora and subcorpora according to the frequencies of the variables. Clustering therefore identifies registers that prefer or avoid ambiguity. By default settings, two clusters were identified. The silhouette measure of cohesion and separation of clustering was 0.5, fair. The variable Subjunctive mood demonstrated the lowest coefficient of importance for clustering (0.08) and was abandoned from the research. Continuing the procedure with five variables, the number of clusters was increased step by step to achieve the highest silhouette measure of cohesion and separation of clustering. The highest value (0.8) was achieved with 17 clusters, but this number was considered excessive. A lower silhouette measure of cohesion and separation of clustering (0.7) was achieved with seven clusters and this number was considered appropriate for analysis (Table 1). Cascades of nouns in the genitive case have the highest coefficient of importance in seven clusters (1.0), followed by NMZ (0.91), PV (0.72), SDP (0.49), and IDP (0.35). The clustering results are visualized in Figure 1. In the image, boxplots

represent the locality, spread, and skewness groups of the data through quartiles; the outlier corpora are identified by circles.

Table 1. The sampled corpora and subcorpora.

Clusters, corpora and subcorpora	Words, mln	Latvian name of the corpora
Cluster 1. Laws		
Laws, 1993-1995	2.5	Likumi
Laws, 2003-2006	11.4	Likumi
Laws, 2019-2022	9.7	Likumi
Cluster 2. Science		
PhD Social sciences, 1993-2020	6.7	Disertācijas
PhD Natural sciences, 1993-2020	1.6	Disertācijas
Cluster 3. Internet media, Parliament, Science		
Internet media, 2001-2021 (split into 10 files)	348.4	Ziņas
National press, 2019-2022	40.0	LVK2022
Regional press, 2019-2022	20.6	LVK2022
PhD Philology, 1993-2020	1.1	Disertācijas
PhD Philosophy, Arts, Arts, 1993-2020	2.1	Disertācijas
Parliament, 1993-1995	2.8	Saeima
Parliament, 2002-2006	1.9	Saeima
Parliament, 2014-2018	1.9	Saeima
Cluster 4. Soviet newspapers		
Cīņa, 1944-1991	184.9	Cīņa
Padomju Jaunatne, 1944-1989	138.1	Jaunatne
Literatūra un Māksla, 1945-1995	52.7	LitMāksla
Rīgas Balss, 1957-1990	98.2	Rīgas Balss
Cluster 5. Essays		
Blogs, 2001-2015	6.7	Emuāri
Pandemia diary, 2020-2022	0.6	PanDi
School essays, 2018	0.2	Pārspringumi
Subtitles of documentary, 2015-2020	5.0	Subtitri
Rainis political essays, 1895-1929	0.2	Rainis

Cluster 6. Conversations

Spoken language, 2012-2024	0.3	LATE-sarunas
Subtitles of fiction films, 2015-2020	0.3	Subtitri
Bērnība (children's magazine), 1945-1964	4.6	Bērnība
Karogs (literary magazine), 1940-1994	48.7	Karogs
Rainis poetry, 1895-1929	0.1	Rainis
Poetry, 1920-1999	3.1	Latviešu 20. gs. dzejas korpus

Cluster 7. Literature

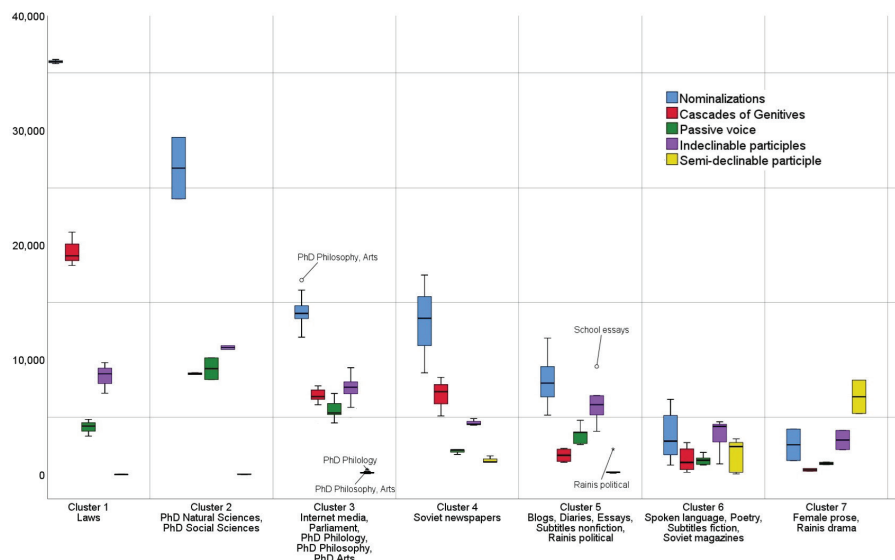
Female prose, 1893-2002	0.9	ĪsprozaS
Rainis drama, 1895-1929	0.3	Rainis

The semi-declinable participle is a marked characteristic of cultivated language use. It is a distinctive feature of literary works: female prose, drama, and poetry. Abundantly, Rainis used the SDP in his political essays. The popularity of SDP in Soviet magazines and newspapers can be explained by the practices of thorough text editing. Soviet editorial offices hired linguists whose tasks included, in addition to correction of orthography and grammar, polishing of expression style. Note also that the subcorpus of doctoral dissertations in philology is the outlier in Cluster 3. The z-test confirms that students of linguistics and literature use the SDP significantly more frequently ($z = 3.5077, p < .001$) than the authors of the second outlier in this cluster – students of philosophy, musicology, and fine arts. The latter, in their turn, use the SDP twice as often as journalists of the internet news media. SDP is not a special characteristic of spoken Latvian. Oral conversations and subtitles of fiction films (scripted dialogues) have the lowest incidence of SDP among the corpora included in Cluster 6. Pearson's bivariate correlation demonstrates moderate negative correlations of SDP with all four AGF (Table 2).

Nominalization is the most popular AGF. Clusters 1 and 2 display the highest frequencies of NMZ: Laws and PhD theses in social sciences and natural sciences. The least frequently NMZ is used in literature, spoken language, Soviet magazines, subtitles, as well as in text written by nonprofessional authors (blogs, diaries, school essays), and Rainis' political essays. The midlevel is occupied by Soviet newspapers and contemporary internet media, parliamentary debates, and dissertations in philology, philosophy, and arts.

The highest incidence of G3+ cascades was found in laws. The score of the followers is almost three times lower: PhD theses, internet media, parliamentary

Figure 1. Frequencies of grammatical forms by clusters, occurrences per mln words.



debates, and Soviet newspapers. The lowest frequencies of the cascades are observed in prose and drama.

Laws, doctoral dissertations, internet media, parliamentary debates, blogs, diaries, and school essays score the highest numbers of IDP. A strong correlation between NMZ and IDP ($r = .666$, $p < .001$, Table 2) requires a special explanation. Using IDP in the subordinate clause, the sender suggests that the action is causally related to the action of the main clause. When the main action is identified by NMZ (see Proposition 4), the causality is difficult, if not impossible, to understand. The combined use of NMZ and IDP creates not only lexical ambiguity concealing perfective (result) and imperfective (process) aspects of an action, but also structural ambiguity because there is no explicit distinction between the main and secondary action. The sender leaves the impression of reasoning about the dynamic processes, but does not assert their actuality and causal necessity.

The use of PV is a characteristic of science, internet media, and law. Non-professional authors of essays tend to use PV less often. The lowest incidence of PV was found in literary works, spoken language, and Soviet magazines and newspapers.

Pearson's bivariate correlation shows that four AGF have moderate to strong mutually positive correlations. The very strong correlation between NMZ and G3+ is

not surprising, because, tagged as nouns, NMZ in the genitive case also appear in cascades.

Table 2. Pearson’s bivariate correlation of the incidence of grammatical forms.

	SDP	NMZ	G3+	PV	IDP
SDP	–	-.470**	-.432**	-.567**	-.516**
NMZ	-.470**	–	.944**	.552**	.666**
G3+	-.432**	.944**	–	.412**	.529**
PV	-.567**	.552**	.412**	–	.745**
IDP	-.516**	.666**	.529**	.745**	–

** . Correlation is significant at the 0.01 level (2-tailed).

Ambiguous objectivity Quantitative analysis rejects the statistical null hypothesis: The difference in relative frequency of ambiguous grammatical forms in the specialized Latvian language corpora is statistically significant. The highest frequency of AGF is observed in law and doctoral dissertations in the natural and social sciences. They are followed by the internet media, dissertations on philosophy, arts, music, and philology, parliamentary debates, and Soviet newspapers. The lowest incidence of AGF was found in literary works, Soviet magazines, subtitles of fiction films, and transcripts of oral conversations. Thus, registers that issue orders, strive for objectivity and impersonality, and involve monological communication tend to have a higher rate of AGF. Registers that admit subjectivity and involve dialogical communication tend to have the lowest AGF rate. The non-ambiguous control variable, SDP, is found more frequently in registers that admit subjectivity and involve dialogue.

The language of Cluster 5 can be explained keeping in mind that these corpora represent a transition point between private and public communication. The writers of blogs, diaries, and essays share their personal experience and opinion, and this down-to-earth routine could have been verbalized in plain language; however, they prefer less clear grammatical forms. I assume that sharing their subjective thoughts with a large anonymous public (any internet user who found the blog), an authority (schoolteachers who evaluate the exam essays), or a significant other (a person to whom the author entrusted her diary), these nonprofessional writers are concerned with finding the appropriate style of public communication. Today, the pervasive internet news media are those institutions that cultivate the linguistic norm of public communication. Note that mass media texts make 60% of the balanced corpus of Latvian, LVK 2022.

The high incidence of AGF in discourses that strive for objectivity and impersonality makes us think that these grammatical forms better grasp complex scientific, legal, and political phenomena, and the authors provide the necessary cues for their disambiguation. The following examples randomly found in dissertations in sociology and political science demonstrate that at least sometimes it is not true. Proposition (8) with five nouns in the genitive case does not create structural ambiguity, as it is easy to split into meaningful units, whereas its lexical ambiguity is high.

- (8) [Konstruktīvisms] sekmē (IND) novērtējuma (NMZ, GEN) pētījuma (GEN) rezultātu (GEN) izmantošanas (NMZ, GEN) prakšu (GEN) attīstību (NMZ, ACC).
[Constructivism] promotes (IND) the development (NMZ, ACC) of practices (GEN) of use (NMZ, GEN) of results (GEN) of evaluation (NMZ, GEN) research (GEN).

Arguably, the author wants policy makers to consider the results of evaluation research when drafting laws and policy documents. In the sentence, three nouns in the genitive case that name the object 'evaluation research results' become a characteristic of the purpose 'to use', but this verb itself is nominalized and put in the genitive case as a specification of another purpose 'practices'; again, used in the genitive case, the noun 'practices' becomes a characteristic of another purpose named by the NMZ 'development' which, in its turn, is devoid of concrete meaning. The author has changed the focus from concrete action 'to use' (policy makers consider research data) to ambiguous 'development'. The Latvian language dictionary Tezaurs.lv defines 'development' as 'qualitative changes'. Since the proposition does not specify the object to be changed and the scale of change, the reader only learns that someone changes something in 'practices' to achieve some new quality. Preceded by the nouns in the genitive case that function as characteristics, the nominalization 'development' appears as a complex scholarly notion, which, in fact, is not.

Next example is a one sentence summary of a scholarly publication.

- (9) [Pētnieki] nonāca (IND) pie (PREP) secinājuma (NMZ),
[Researchers] came (IND) to (PREP) the conclusion (NMZ),
(10) ka (CONJ) arodbiedrību (GEN) attīstībā (NMZ, LOC) turpinās (IND) arodbiedrību (GEN) īpatsvara (GEN) samazināšanās (NMZ, NOM),
that in the development (NMZ, LOC) of trade unions (GEN) is continuing (IND) a decrease (NMZ, NOM) of the share (GEN) of trade unions (GEN),
(11) pārrunu (GEN) starp (PREP) darba (GEN) devējiem (DAT) un (CONJ) arodbiedrībām (DAT) decentralizācija (NMZ, NOM),
the decentralization (NMZ, NOM) of negotiations (GEN) between (PREP) work (GEN) givers (DAT) and (CONJ) trade unions (DAT),
(12) kā arī (CONJ) konstatēja (IND) darbinieku (GEN) līdzdalības (NMZ, GEN) nozīmes (NMZ, GEN) samazināšanās (NMZ, GEN) tendences (ACC).
as well as (CONJ) found (IND) the trends (ACC) of decrease (NMZ, GEN) of significance (NMZ, GEN) of participation (NMZ, GEN) of employees (GEN).

In the sentence, there is a collective human agent ('researchers') and three verbs in indicative mood (researchers 'came' and 'found', the decrease 'is continuing'). Proposition (9) raises no interpretation problem: 'came to the conclusion' means 'discovered'. Subsequent ambiguity is due to four nominalizations and two cascades of the genitives. The NMZ 'development' in (10) is not only ambiguous like in (8), but it is also illogical: How can the share of trade unions decrease in the development of trade unions? NMZ 'decentralization' in (11) conceals the object and the agent: Where is the 'center' of negotiations and who decentralizes them? NMZ 'decrease' conceals agents in (12): Who sidelines employees? The cascade 'significance of participation of employees' conceals the agent and object: What is the 'significance' and for whom? Proposition (12) can be understood in various ways:

- (13) Employees are unwilling (or unable, distrustful) and therefore they do not participate in trade unions (or in negotiations with employers).
- (14) Employers (or trade union leadership) are unwilling (or unable, distrustful) and therefore employees are not admitted to participation.

Nominalizations and cascades of the genitives conceal much of the information necessary to reformulate the sentence in concise language. Based on personal knowledge, I would rewrite the sentence: 'Someone distrusts trade unions; employers negotiate with someone else directly, but not with trade unions; employees do not want to participate in negotiations with employers'. The author can always refute my reading as inappropriate because the text does not explicitly identify the agents and their actions.

Conclusion

The difference in the relative frequency of ambiguous grammatical forms between Latvian language corpora is statistically significant. Most often, they are used by powerful social institutions: the state, science, and the internet news media. To disambiguate meaning, receivers look after the interpretation cues, but not always senders include them in the cotext and explicate the relevant context. AGF creates an impression of objective and impersonal reasoning about complex reality, but in fact the sender shifts responsibility for communication failure onto the receiver, who is not able to identify subject, object, and predicate of the proposition and to understand the implied causality. In this way, ambiguity maintains hierarchical relations between the sender and the receiver.

The ambiguity of grammatical forms is a common feature of natural languages. There is no reason to avoid them altogether, because the cotextual and contextual information helps communicators to disambiguate meaning. The extent to which

senders misuse ambiguity for manipulative purposes can be established by a combined qualitative and quantitative analysis of a random sample of laws, doctoral dissertations, and internet news.

List of abbreviations

ACC	Accusative case
AGF	Ambiguous grammatical forms
CONJ	Conjunction
DAT	Dative case
DPPP	Declinable present passive participle
G3+	Cascade of three or more nouns in the genitive case
GEN	Genitive case
IDP	Indeclinable participle
IND	Indicative mood
LOC	Locative case
NMZ	Nominalization
NOM	Nominative case
PART	Particle
PREP	Preposition
PV	Passive voice
SDP	Semi-declinable participle
SUB	Subjunctive mood

References

- Fairclough, Norman (2013). *Language and Power*. London: Routledge.
- Fowler, Roger (1991). *Language in the News*. London: Routledge.
- Fowler Roger; Kress, Gunther (1979) Critical linguistics. R. Fowler, R. Hodge, G. Kress, T. Trew (eds.). *Language and Control*. London: Routledge & Kegan Paul, pp. 185–213.
- Gillon, Brendan S. (1990). Ambiguity, Generality, and Indeterminacy: Tests and Definitions. *Synthese*, No. 85, pp. 391–416.
- Habermas, Jürgen (1998). *On the pragmatics of communication*. MIT Press.
- Kalnača, Andra; Lokmane, Ilze (2021). *Latvian Grammar*. Rīga: University of Latvia Press.
- Kennedy, Christopher (2011). Ambiguity and Vagueness. C. Maienborn, K. von Heusinger, P. Portner (eds.). *Semantics*. Berlin: DeGruyter, pp. 507–535.

Kruks, Sergejs (2023). Ambiguous grammar of legal discourse. *Letonica*, No. 53, pp. 168–185.
DOI: 10.35539/LTNC.2023.0053.09

Nītiņa, Daina; Grigorjevs, Juris (eds.) (2013). *Latviešu valodas gramatika*.
Rīga: Latvijas Universitātes Latviešu valodas institūts.

Titscher, Stefan; Meyer, Michael; Wodak, Ruth; Vetter, Eva (2000). *Methods of text and discourse analysis*. London: SAGE.