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.2023.0053.09

Ambiguous Grammar of Legal Discourse

Neviennozīmīgas gramatiskās formas juridiskajā diskursā

Keywords:
Latvian language,
corpus analysis,
statistics of language use,
factor analysis of language use,
ambiguous grammatical form

Atslēgvārdi:
latviešu valoda,
korpusa analīze,
valodas lietojuma statistika,
valodas lietojuma faktoranalīze,
neviennozīmīga gramatiskā forma

Summary A factor analysis of the corpus of Latvian legal documents issued during 1990–2022 reveals an increase of the use of ambiguous grammatical forms that obscure the regulatory enactments. Designating actions and their effects by indeclinable participles and verbs in conditional mood, the lawmaker does not attest to causality between the proposed procedure and the result sought, and does not acknowledge their intention to implement the procedure. The indicative mood and the infinitive forms designate actions and their effects explicitly, but their frequency tends to decrease. The frequency of usage of passive voice, debitive mood, and reflexive verbs demonstrates an upward trend; however, assessing the ambiguity of these grammatical forms requires a qualitative analysis of legal documents.

Kopsavilkums Latvijas likumu korpusa (1990–2022) faktoranalīze apliecina, ka pieaug neviennozīmīgu gramatisko formu lietojums, kas normatīvo aktu saturu padara neskaidru. Apzīmējot darbības un to rezultātus ar nelokāmiem divdabjiem un darbības vārdiem vēlējuma izteiksmē, likumdevējs negarantē cēloņsakarību starp piedāvāto rīcību un vēlamo rezultātu, kā arī neapliecina gatavību īstenot iecerēto. Darbības un to rezultātus precīzi apzīmē darbības vārdi īstenības izteiksmē un nenoteiksmē, taču šo gramatisko formu lietojums pakāpeniski samazinās. Ciešamās kārtas, vajadzības izteiksmes un atgriezenisko darbības vārdu lietojuma biežumam ir tendence pieaugt, taču šo gramatisko formu ietekme uz juridiskā diskursa neviennozīmību jāpārbauda ar kvalitatīvo metodi.

Introduction

Analyzing Latvian policy documents, I have noticed a growing frequency of indeclinable participles during the last few years (Kruks 2020). Combining the characteristics of verb and adverb, the indeclinable participles increase the ambiguity of meaning. In some cases, even the context does not allow the reader to determine whether the proposition describes conditions for an action, an action unfolding over time, or an already accomplished action. Mediated by government press releases, this grammatical form has also become prominent in the news media. Is it merely a subjective observation or a statistically significant trend? This article analyzes the diachronic dynamics of grammatical forms in Latvian legal documents issued during 1990–2022. The first section discusses ambiguous grammatical forms and the social implications of their use. The second section describes the methodology of data collection and statistical analysis. The third section interprets the data.

Ambiguous grammar

Ambiguity is a kind of uncertainty "that manifests itself as a variation in truth conditions: one and the same utterance token can be judged true of one situation and false of another, or the other way around, depending on how it is interpreted" (Kennedy 2011: 508). Ambiguity is common in oral communication because the context can provide enough information for understanding the message. Steven Piantadosi, Harry Tily, and Edward Gibson (2012) even consider ambiguity as a criterion for an efficient communication system: the verbal message does not convey information already provided by the context. They mention a dozen research articles that demonstrate the ability of receivers to use contextual information in the form of discourse context, local linguistic context, or more global knowledge about world in disambiguating language. Contextual interpretation is a necessary feature of the legal language (Hart 2012, Endicott 2000, Tiersma 1999). There might be situations where a statute cannot or should not be applied straightforwardly: ambiguous formulations give judges a chance and even require them to consider specific contextual circumstances not covered explicitly in the law.

There are two types of linguistic ambiguity (Gillon 1990; Kennedy 2011). Lexical ambiguity occurs when a word has more than one meaning; structural ambiguity emerges when a phrase can be segmented in more than one way. This research is not concerned with the use of terms that denote different phenomena (for example, 'nation' as a political or cultural concept), but with the ambiguity of grammatical

forms. In some grammatical forms, a word loses the typical characteristics of its part of speech or acquires characteristics of other parts of speech. In order to interpret the meaning, the reader has to decide which part of speech they are dealing with now. When we speak and write, we select words and grammatical forms from a wide range of linguistic resources. Critical approaches to language studies suggest that forms of language encode a socially constructed representation of the world; the author may intentionally select a specific grammatical form to conceal some aspects of reality (Fowler 1991; Halliday 1970; Hodge and Kress 1993; Pêcheux 1975). The researcher's task is not to interpret the meaning, but to draw the reader's attention to unclear strategic intentions. The critical approach in Latvian linguistics is rare. The encyclopedic 475-page book *The Latvian Language,* written by 20 famous linguists, does not touch upon the issue at all (Veisbergs 2013). Daina Nītiņa (2001) as well as Andra Kalnača and Ilze Lokmane (2021) have identified some grammatical forms that can carry several meanings at once, but the purpose of their publications was beyond a critical analysis of the use of ambiguous language in communication. The following list of grammatical forms, which are the focus of the present research, is an adaptation of English and French language studies to Latvian language. A common effect of ambiguity is that it allows speakers – in our case, public governance institutions issuing legal acts – to eschew responsibility. From their ambiguous statements, it is not clear who is undertaking, will undertake, or must undertake an action, and whether this action will lead to the declared goal.

The most discussed grammatical form is the **passive voice** of verbs. Concealing the agent, the passive forms dissimulate the causality of action, presenting the result as an effect of objective forces beyond human control. In Latvian, the indefinite passive tense forms are derived by means of the auxiliary verb *tikt* ('to get') in a finite form plus past passive participle; the perfect passive tense forms use the auxiliary verb *būt* ('to be') in a finite form plus past passive participle (Kalnača, Lokmane 2021: 260). In sentences, passive voice forms express generalized, regular or habitual actions, or impersonal statements of a fact (Kalnača, Lokmane 2021: 263). This variety of meaning allows speakers to present their subjective opinion as a generalization or a fact.

There are two types of **reflexive verbs**, and one of these types does not pose problems for interpretation. These are subject reflexive verbs that designate actions involving the subject's body or a reciprocal interaction between subjects. Ambiguity pertains to the object reflexive verbs: in propositions, the patient turns into a syntactic subject, while the agent acting upon the patient is dissimulated (Kalnača, Lokmane 2021: 272–281). A qualitative research of Latvian language parliamentary discourse revealed that decision-makers most commonly used object reflexive verbs and passive voice to dissimulate the agent in their speech (Ījabs, Kruks 2008).

Nominalizations are substantiated verbs. As nouns, they name actions while concealing the categories of the verbs from which they are derived: person, number, tense, and mood. Verbs and verb phrases are more basic than certain classes of nouns; anything that makes a verb less verb-like and more noun-like creates abstractions (Charrows, Charrows 1979: 1321). By replacing entire subordinate clauses, nominalizations eliminate the agent, thus making the sentence vague, impersonal, and hard to reconstruct. Nominalizations suppress the narrative structure of the message, which is the basic form of describing and understanding the world. This research is concerned with those nominalizations which denote processes rather than concepts, places, machines etc. In Latvian these nouns are derived from verbs by suffixes *-šan-* and *-īb-* (Kalnača, Lokmane 2021: 131–137).

Cascades of genitive have been studied by Patric Sériot (1986) when analyzing the Russian official language. In Latvian, the phenomenon is even more widespread (though no comparative statistical analysis has been done yet) because adjectivity is often expressed by nouns in the genitive case. For example, a common notion in policy documents *Latvijas kultūras identitāte* can be translated as 'Latvia's cultural identity' or 'identity of the culture of Latvia'. Interpretation depends on the segmentation of the expression, and the segmentation on its turn hinges on the world knowledge. In the example cited above, even this knowledge is not helpful, because both interpretations can constitute the subject of policy documents.

The indeclinable participle combines the characteristics of adverb and verb by expressing mode, time, cause, purpose, and circumstances of (or conditions for) an action (Nītiņa, Grigorjevs 2013: 585). Like the cascades of genitive, sometimes the meaning of an indeclinable participle cannot be interpreted even knowing the logic of the social world. This is how the National Development Plan states the government's goal for 2021–2027: *Efektīva pierobežas ekonomiskā potenciāla izmantošana* [NOMINALIZATION], *stiprinot* [INDECLINABLE PARTICIPLE] *sadarbības* [NOMINALIZATION] *saišu veidošanu* [NOMINALIZATION] ('Effective use of border-zone economic potential, strengthening the creation of cooperation bonds'). Is the effective use a precondition of cooperation bonds or vice versa; does the sentence state a fact or express an intention or a commitment to attain a goal by the year 2027? Surrounded by nominalizations, the indeclinable participle becomes even more difficult to understand logically, because nominalizations as abstract nouns do not refer to concrete real-life phenomena.

The purpose of this research is not to strike a balance between contextual flexibility and certainty of interpretation. If ambiguity is a necessary feature of legal language, then the use of ambiguous grammatical forms should not demonstrate a statistically significant variability over time; this is the statistical null hypothesis of the study.

Method The hypothesis will be tested by using a Corpus of Legal Acts of the Republic of Latvia that contains documents issued by the Parliament, the Government, and the local governments during 1990–2022. The collection of Latvian corpora at *https://korpuss.lv/* enables a qualitative analysis in the *NoSketch Engine*. The extraction of diachronic quantitative data is a rather time-consuming process in this programming environment because the grammatical forms must be counted in many sets of selected documents separately and in order. The *Linux Bash* programme enables quick processing of all sets at once; besides, the extracted statistical data are being stored automatically. Lemmatized and tagged documents merged in a single file usable in *Linux* are deposited in the Common Language Resources and Technology Infrastructure (Skadiņa et al. 2022). The Corpus of Legal Acts was downloaded from this repository: https://repository.clarin.lv/repository/xmlui/handle/20.500.12574/65/.

The corpus represents the texts of legal documents in three columns: words, lemmas, and tags; besides it contains metadata for each document. In *Linux Bash*, the corpus was cleaned from metadata; columns of lemmas and tags were copied into a working file that was split into 100 even files containing 974,624 words each. Then 16 linguistic units (various grammatical forms of various parts of speech) were extracted by the respective tags using the *Bash* commands (see Supplement; the list of tags is available at: https://korpuss.lv/static/media/LV_TagSet_v.2.2.1_22092021.pdf):

- Verbs: reflexive verbs, indicative/conditional/debitive mood, infinitive, *tikt* as the auxiliary verb of the simple tenses of passive voice;
- Participles: declinable and indeclinable;
- Nouns: nominative, genitive, dative, accusative, and locative case forms of common nouns;
- Adjectives;
- Adverbs;
- Nominalizations.

Nominalizations are tagged as nouns in the corpus. For research purposes, nominalizations ending with the most productive suffix *-šan-* were extracted from the list of lemmas. Since the suffix *-īb-* is not unique for nominalizations, and extraction of these words would have required manual work, they were not analyzed. Arguably, nominalizations with *-šan-* would suffice to reveal the diachronic trend, while counting their absolute frequencies is not the purpose of this study.

Pronouns, prepositions, numerals, conjunctions, interjections, particles, and abbreviations were not used in the statistical analysis. Verbs in relative and imperative mood and semi-declinable participles were omitted due to their low frequency: less than 1000 occurrences per file.

Results Statistical analysis of the data extracted was performed in the R programme. Principal Component Analysis was applied to reduce the large amount of data to a smaller number of correlated factors that would have revealed the theoretical construct that helps interpretation of the data. Computation of the Eigenvalues determined five components with quality scores > 1. The result suggests that five components underlie the research problem; therefore, the number of factors to be extracted was set at five. Kaiser-Meyer-Olkin factor adequacy returned a good result, KMO = 0.64. A suggested cut-off point to determine the factorability of the sample data is KMO ≥ 0.60. Bartlett's Test of Sphericity confirms the statistical significance, χ^2 (120) = 927.305, $p < .001$. This means that the variables are not essentially different from one another; their mutual correlation creates factors. Test of the hypothesis that five factors are sufficient returned a statistically significant result: χ^2 (50) = 112.52, $p < .001$. Cumulatively, five factors explain 67% of the variation. The extracted factors with a correlation coefficient greater than 0.50 are summarized in the Rotated component matrix (Table 1).

Table 1. Rotated component matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Conditional mood	0.85				
Indeclinable participle	0.74				
Adjective	0.70				
Declinable participle	0.59				
Adverb	0.58				
Locative case		0.75			
Genitive case		0.65			
Nominalization		0.57			
Nominative case		–.0.66			
Debitive mood			0.76		
Auxiliary verb *tikt*			0.75		
Reflexive verb			0.72		0.51
Dative case			0.52		
Infinitive				0.89	
Accusative case				0.69	
Indicative mood					0.96

Extraction method: Principal Component Analysis

Rotation method: Varimax rotation with Kaiser normalization

Table 2 Reliability of the factors

Factor	Cronbach's α
1	0.644
2	– 0.093
3	0.648
4	0.741
5	0.363

The Principal Component Analysis suggests the validity of the established factors, and their reliability is verified by computing Cronbach's α values (Table 2). Usually, values α ≥ 0.70 are treated as acceptable but in some circumstances factors with lower values of α can be useful (Schmitt 1996). Some researchers recognize values α ≥ 0.64 as "adequate" (Taber 2017). In our case, Factors 2 and 5 do not even meet the lowered reliability requirement. However, in order to compare the dynamics of verb use, a component of Factor 5 – indicative mood – was retained as a unique independent variable. The items of Factor 2 were independently analyzed to assess their investment in diachronic variation. For the sake of comparison, three factors with the adequate Cronbach's α level were remodelled as scales and the values of the scales were normalized using the Min-Max Scaling method. The normalized data values fit in a range between 0 and 1 only, and the effect of outliers on the data values is suppressed to a certain extent in the scales. To avoid confusion, the scales and independent items will be referred to as the numbered factors listed in Table 1.

Discussion and conclusion

Factor 1 includes adjectives, adverbs, declinable and indeclinable participles, and verbs in conditional mood. What are their functions? Adjectives describe the properties of objects, whereas adverbs characterize actions, properties, circumstances and, less frequently, objects (Kalnača, Lokmane 2021: 148, 316). Declinable participles fulfil various functions (Kalnača, Lokmane 2021: 292–296). Declinable present active participles are attributives in sentences. Declinable past active participles sometimes function as attributives but more often as predicates. Declinable present passive participles are predicates; they can express a possibility, necessity, and conditions; they can also function as attributives. Declinable past passive participles denote accomplished actions; they can function as a part of passive voice forms or as adverbs. Despite the variety of functions, there is no evidence that the declinable participles increase ambiguity of a

proposition, but in some forms they explicitly denote the characteristics of an action rather than the action itself. As to the indeclinable participles, their ambiguity owes to a confusion of functions that are difficult to discern even in the pragmatic context (discussed in Section 1). Some indeclinable participles function as explanations. In legal acts, they mediate intertextuality – for example, *atsaucoties uz likumu* ('referring to the law') or *ievērojot Ministru kabineta lēmumu* ('following the Government decision'). The problem with the meaning of these explanations is twofold. If this is a common feature of legal language, its use should not demonstrate statistically significant variability. On the other hand, intertextual reference to an entire document results in ambiguity if a concrete utterance of the antecedent document is not specified and its pertinence to the new document is not explicated.

The main interest of this study was in the five grammatical forms described in Section 1, but factor analysis suggests examining conditional mood as another source of ambiguity. In Latvian language this feature of the verb is called 'optative mood' (*vēlējuma izteiksme*), which corresponds to the English term 'conditional mood' (Skujiņa 2007: 440), whereas English linguists also use the term 'subjunctive mood' to designate the phenomenon. Mood is a grammatical category coding modality which is the abstract semantic functional category that refers to the speaker's attitude to the proposition (Becker, Renberger 2010: 1). The speaker may express assertion or non-assertion of the proposition, or indicate that the proposition is about factual or non-factual events and states (Palmer 2001: 1–3). There are several types of modality. Epistemic modality refers to the speaker's judgment of the truth of the proposition (Palmer 1990: 6) or the speaker's commitment to the object of the proposition (Verstraete 2001: 1517; Nuyts 2016b). Non-epistemic modalities are "future-projecting" (Ziegeler 2003, 2006) because they refer mostly to goal-centred situations (Frawley 1992: 425), or antecedents to action (James 1986: 51), or purpose (van Olmen, Auwera 2016: 368). Future-projecting propositions require modal expression because the speaker cannot be fully factual about events that have not occurred yet; the propositions without modality, in their turn, connote more certainty. Differences in modality use are related to the speaker's knowledge about the subject of the proposition (Huddleston, Pullum 2005). As to the choice of morphological or lexical markers of modality, languages vary in deciding what is factual and what is not (Bybee et al. 1994; Givón 1994). This is evident even in translations of straightforward legal documents. Let us take, for example, the English original of this European Union document and its official translation into Latvian language:

> In order **to promote** the sustainable management of forest resources, the Parties commit to work together **to improve** forest law enforcement and governance and **to promote** trade in legal and sustainable forest products…

> Lai **veicinātu** [CONDITIONAL] meža resursu ilgtspējīgu pārvaldību, puses apņemas sadarboties, lai **uzlabotu** [CONDITIONAL] meža tiesību aktu ieviešanu un pārvaldību un **veicinātu** [CONDITIONAL] likumīgu un ilgtspējīgu mežsaimniecības ražojumu tirdzniecību…
>
> (Comprehensive Association Agreement between Central America and the European Union. December 15, 2012. Article 289. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2012.346.01.0003.01.ENG&toc=OJ%3AL%3A2012%3A346%3ATOC/)

In English, the purpose is expressed by verb infinitives that identify actions. The Latvian translator has preferred conditional mood. Latvian grammar prescribes this verb form in the purpose clauses introduced by the conjunction *lai* ('in order to'), but in addition the same form shows that the action is desired or possible under certain conditions (Nītiņa, Grigorjevs 2013: 496). Debra Ziegeler points out that the distinctions between different types of modality are not formally marked in English, and "it is only the semantic and pragmatic context in which the construction is uttered that determines which applies" (Ziegeler 2019: 427). The same is true for Latvian. In contrast to the original text with infinitives, the Latvian text may be interpreted by the reader that the legislator is not sure of the achievement of the goal. In another document, the Latvian legislator has avoided the ambiguity by using indicative mood to identify a goal:

> [Lauksaimniecības ministrija] izstrādā ekonomiskos un tiesiskos noteikumus, lai novērstu [CONDITIONAL] atsevišķu uzņēmumu un apvienību monopolstāvokli un veicinātu [CONDITIONAL] nozares iekšējo konkurenci; organizē un stimulē [..] tādu tehnoloģisko procesu izstrādi un ieviešanu, kas aizsargā [3rd PERSON, PRESENT, INDICATIVE] dabas resursus. (LR Ministru Padomes lēmums Nr. 80 Par Latvijas Republikas Lauksaimniecības ministriju. Decision No 80 of the Council of Ministers 'On the Ministry of Agriculture of the Republic of Latvia. Article 2. July 31, 1990)
>
> [Ministry of Agriculture] develops economic and legal regulations in order **to prevent** [that would prevent] the monopoly status of some companies and associations and **to promote** [that would promote] internal competition in the industry; organizes and stimulates [..] the development and implementation of technological processes **that protect** natural resources.

The indicative mood of the third purpose expresses a deontic modality: the government demands that the ministry foster technologies that *actually do protect* natural resources. On this background, the conditional mood of the first two purposes lacks assertive force: the statements express merely a wish. Nina Dobrushina et al. remind that an expression of a wish is not equal to an attainment of a goal: "With the optative, the state of affairs wished for is typically outside the sphere of influence of the speaker" (2005: 299). Conditional mood connotes epistemic modality here: the government evaluates the likelihood that the ministry is in capacity to edit adequate regulations and that these regulations indeed will transform the reality.

Legal discourse projects into the future, and the inevitable non-factuality calls for conditional mood. At the same time, grammatically correct propositions increase ambiguity because a verb in conditional mood designates different modalities that are not distinguished by morphological markers. Note that adjectives and adverbs included in this factor also function as lexical devices of modality (Nuyts 2016a). In official documents, modality can be read as a declaration of uncertainty and lack of commitment, rather than a grammatical device of a future-projecting proposition.

Factor 2 that includes the grammatical forms of nouns does not pass the reliability test. The exclusion of some items from the scale increases reliability, but the maximum value of Cronbach's α is still unsatisfactory, 0.431. Regression analysis of items reveals that the nouns in the nominative case have the strongest effect on diachronic variability: standardised coefficient $\beta = -.497$, $p < .001$. The effect of nouns in the locative case is less strong: $\beta = .236$, $p = .022$. As the frequency of nouns in the nominative case decreases, the locative case is used more often. The frequency of nouns in the genitive case does not vary statistically significantly. The data reflect the trend of individual occurrences of genitive; in order to assess frequency of the cascades of genitive, a different computation method is needed. Neither does the frequency of nominalizations ending in *-šana* experience a statistically significant variation.

The grammatical forms included in Factor 3 do not necessarily increase ambiguity. A higher potential of ambiguity pertains to object reflexive verbs, but in corpus statistics they are not distinguished from subject reflexive verbs. Passive voice allows generalizations that do not need to specify agents; in this capacity, it is appropriate in legal discourse. In the English legal language, passives do not create much confusion when used in main clauses, whereas in dependent clauses they are more difficult to understand (Charrow, Charrow 1991:1325). The debitive mood expresses a necessary or required action (Nītiņa, Grigorjevs 2013: 497). In this capacity, it is appropriate in legal discourse which instructs individuals and institutions to undertake or not to undertake certain actions. Ambiguity arises when verbs in debitive mood are used in impersonal sentences, but the current quantitative methodology does not distinguish such cases. The use of reflexive verbs in debitive mood is not an important source of ambiguity: 5% of verbs in debitive mood are used in this grammatical form.

A noun in dative case indicates something/somebody for whom something is intended or at whom something is directed, or which has/possesses the concept expressed by a verb or noun (Nītiņa, Grigorjevs 2013: 350). Inclusion of dative in the factor is logical because it is most often used with verbs. Statistics corroborates this observation: the use of nouns in the dative case correlates with the use of verbs in debitive mood ($r = .410$) and passive verbs ($r = .456$). These are the highest correlations for the dative case in the corpus.

Factor 4 includes verbs in the infinitive and nouns in the accusative case. Infinitives name the action itself, whereas nouns in the accusative case denote objects to which the action refers. Infinitive forms identify actions in a general way, they can also express a necessity or need (Nītiņa, Grigorjevs 2013: 460, 462). The independent variable of Factor 5 – indicative mood – refers to an action that is actually taking place, has taken place, or will take place (Nītiņa, Grigorjevs 2013: 494). Researchers of the English legal language suggest using verb infinitives because they underline the action that is regulated by the document; indicative mood enables narration of events, and is more often used in contracts which explicitly designate agents and actions (e.g., Bix 1993; Tiersma 1999).

Graphic representation of the relative change in the frequency of grammatical forms during 1990–2022 attests the diachronic dynamic of language use (Figure 1). Several sharp peaks point to documents that are not in general trend, thus making the graph difficult to read. Locally estimated scatterplot smoothing, LOESS, provides a more comprehensible visualization of the diachronic dynamics. LOESS is a non-parametric method for smoothing a series of data, in which no assumptions are made about the underlying structure of the data. Local regression is used to fit a smooth curve through a scatterplot of data (Cleveland 1979). The statistically significant result of LOESS ($p < .05$) is represented in Figure 2.

In the last quartile of the analyzed time span, Factor 1 gains importance. The main benefit of factor analysis is that the researcher can focus on the unique core elements of the variables instead of their redundant attributes. A common feature of the items of this factor is conditionality – in other words, the items describe the conditions and the speaker's attitudes towards the action. The items of Factor 3 are not necessarily ambiguous per se because they allow for generalizations necessary in legal discourse. The relative proportion of passive voice, debitive mood, and reflexive verbs was the highest at the beginning of the studied period, reached a minimum in the middle of the period, but then a slow increase began. On this background, the proportion of ambiguous grammatical forms in Factor 1 demonstrates a steady growth, and by the fourth quartile of the period this factor explains more variance in language use. In other words, the ambiguity of meaning in the legal discourse is growing. Meanwhile, the proportion of indicative mood and infinitive forms is slightly decreasing.

Quantitative analysis of the corpus of legal documents rejected the statistical null hypothesis. Parts of speech and grammatical forms constitute factors that demonstrate a diachronic dynamic. The proportion of ambiguous grammatical forms has a statistically significant increasing trend. Conditional mood verb forms and indeclinable participles contribute to the ambiguity of legal discourse because they

dissimulate those semantic aspects of verbs that indicate agent and results. The legislator describes a desired state to be attained but is not sure whether the proposed action will lead to the result or whether the responsible agent can attain the stated goal. The proportion of passive voice, debitive mood, and reflexive verbs is also increasing, but the ambiguity of their meaning depends on the pertinence of generalizations that can be established in qualitative analysis.

References

Becker, Martin; Remberger, Eva-Maria (2010). Modality and Mood in Romance: An introduction. M. Becker, E.-M. Remberger (eds.). *Modality and Mood in Romance*. Berlin: De Gruyter, pp. 1–12.

Bix, Brian H. (1993). *Law, Language, and Legal Determinacy*. Oxford: Clarendon Press.

Bybee, Joan L.; Fleischman, Suzanne (eds.) (1995). *Modality in Grammar and Discourse*. Amsterdam/Philadelphia: John Benjamins.

Charrow, Robert P.; Charrow, Veda R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, No. 7, pp. 1306–1374.

Cleveland, William S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* No. 74, pp. 829–836.

Dobrushina, Nina; van der Auwera, Johan; Goussev, Valentin (2005). Optative. M. Haspelmath, M. Dryer, D. Gil, B. Comrie (eds.). *World Atlas of Language Structures*. Oxford: Oxford University Press, pp. 298–301.

Endicott, Timothy A. O. (2000). *Vagueness in Law*. Oxford: Oxford University Press.

Fowler, Roger (1991). *Language in the News: Discourse and Ideology in the Press*. London: Routledge.

Frawley, William (1992). *Linguistic Semantics*. Hillsdale: Lawrence Erlbaum.

Gillon, Brendan S. (1990). Ambiguity, Generality, and Indeterminacy: Tests and Definitions. *Synthese*, No. 85, pp. 391–416.

Givón, Talmy G. (1994). Irrealis and the subjunctive. *Studies in Language*, No. 18, pp. 265–337.

Halliday, Michael A. K. (1970). Language Structure and Language Function. J. Lyons (ed.). *New Horizons in Linguistics*. Harmondsworth: Penguin, pp. 140–165.

Hart, Herbert L. A. (2012). *The Concept of Law*. Oxford: Oxford University Press.

Hodge, Robert; Kress, Gunther (1993). *Language as Ideology*. London: Routledge.

Huddleston, R.; Pullum, G. K. (2005). *A Student's Introduction to English Grammar*. Cambridge: Cambridge University Press.

Ījabs, Ivars; Kruks, Sergejs (2008). *Saeima, vārdi un demokrātija*. Rīga: Sorosa Fonds – Latvija.

James, Francis (1986). *Semantics of the English Subjunctive*. Vancouver: University of British Columbia 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 (2020). Uzticības, sadarbības un vienotības konceptu izpratne Nacionālajā attīstības plānā 2021.–2027. gadam. *Akadēmiskā Dzīve*, No. 56, pp. 131–147.

Nītiņa, Daina (2001). Vārdu un vārdformu semantikas pragmatiskie aspekti [Pragmatic aspects of semantics of words and wordforms]. *Linguistica Lettica*, No. 8, pp. 78–86.

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

Nuyts, Jan (2016a). Surveying modality and mood. An introduction. J. Nuyts, J. van der Auwera (eds.). T*he Oxford Handbook of Modality and Mood.* Oxford: Oxford University Press, pp. 1–8.

Nuyts, Jan (2016b). Analyses of the modal meanings. J. Nuyts, J. van der Auwera (eds.). *The Oxford Handbook of Modality and Mood.* Oxford: Oxford University Press, pp. 31–48.

Nunnally, Jum C.; Bernstein, Ira H. (1994). *Psychometric Theory.* McGraw-Hill.

Palmer, Frank R. (1990). *Modality and the English Modals*, 2nd edition. London: Longman.

Palmer, Frank R. (2001). *Mood and Modality.* Cambridge: Cambridge University Press.

Pêcheux, Michel (1975). *Les Vérités de La Palice.* Paris: Maspero.

Piantadosi, Steven T.; Tily, Harry; Gibson, Edward (2012). The communicative function of ambiguity in language. *Cognition*, No. 122, pp. 280–291.

Schmitt, Neal (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, No. 8(4), pp. 350–353.

Sériot, Patrick (1986). How to Do Sentences with Nouns: Analyzing Nominalizations in Soviet Political Discourse. *Russian Linguistics*, No. 10(1), pp. 33–52.

Skadiņa, Inguna; Auziņa, Ilze; Darģis, Roberts; Voitkāns, Arnis (2022). CLARIN valodas resursu un rīku pētniecības infrastruktūra humanitārajām un sociālajām zinātnēm. *Letonica*, No. 47, pp. 312–327.

Skujiņa, Valentīna (ed.) (2007). *Valodniecības pamatterminu skaidrojošā vārdnīca.* Rīga: LU Latviešu valodas institūts.

Taber, Keith S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, No. 48(1), pp. 1273–1296.

Tiersma, Peter M. (1999). *Legal Language.* Chicago: University of Chicago Press.

Van Olmen, Daniel; van der Auwera, Johan (2016). Modality and Mood in Standard Average European. J. Nuyts, J. van der Auwera (eds.). *The Oxford Handbook of Modality and Mood.* Oxford: Oxford University Press, pp. 362–385.

Veisbergs, Andrejs (ed.) (2013). *Latviešu valoda* [Latvian language]. Rīga: LU Akadēmiskais apgāds.

Verstraete, Jean-Cristophe (2001). Subjective and objective modality: Interpersonal and ideational functions in the English modal auxiliary system. *Journal of Pragmatics*, No. 33, pp. 1505–1528.

Ziegeler, Debra (2003). On the generic origins of modality in English. D. Hart (ed.). *English Modality in Context: Diachronic Perspectives*. Bern: Peter Lang, pp. 33–69.

Ziegeler, Debra (2006). *Interfaces with English Aspect: Diachronic and Empirical Studies.* Amsterdam: John Benjamins.

Ziegeler, Debra (2019). Mood and Modality. B. Aarts, J. Bowie, G. Popova (eds.). *The Oxford Handbook of English Grammar.* Oxford: Oxford University Press.

Supplement. List of the *Linux Bash* commands

Removing <html codes> from the original file.

sed -E 's/<.*?>//' likumi.txt > likumi2.txt

Removing empty lines.

awk '!/^$/' likumi2.txt > likumi3.txt

Extracting columns of tags and lemmas.

awk '{print $2 " " $3}' likumi3.txt > likumi4.txt

Counting lines / words (each line contains one word).

wc -l likumi4.txt

Splitting the file evenly into 100 files: 974,625 lines/words per file.

awk 'NR%974625 ==1 {x="Law_"++i;} {print > x}' likumi4.txt > Law_1...100.txt

Renaming files Law_1 ... Law_9 to Law_01 ... Law_09 and Law_100 to Law_99b

mv Law_1 Law_01

.... etc.

Extracting tags | removing names of tags | sorting and counting words > storing data in a file. (More tagging info is added in each query to avoid confusion with lemmas, notably there were some abbreviations resembling part of tags).

egrep -o 'v.[ny]pu' Law* | sed 's/:v.*pu//g' | sort | uniq -c > 01PartID.txt

egrep -o 'v.[ny]pd' Law* | sed 's/:v.*pd//g' | sort | uniq -c > 01PartD.txt

egrep -o 'vm[ny]i' Law* | sed 's/:vm[ny]i//g' | sort | uniq -c > 01VerbMainInd.txt

egrep -o 'vc[ny]' Law* | sed 's/:vc[ny]//g' | sort | uniq -c > 01VerbBe.txt

egrep -o 'va[ny].....[ny]' Law* | sed 's/:va[ny].....[ny]//g' | sort | uniq -c > 01VerbTikt.txt

egrep -o 'v.[ny]c[pfs0]' Law* | sed 's/:v.[ny]c[pfs0]//g' | sort | uniq -c > 01VerbCond.txt

egrep -o 'v.[ny]d.....[ny]' Law* | sed 's/:v.[ny]d.....[ny]//g' | sort | uniq -c > 01VerbDebitive.txt

egrep -o 'v.[ny]n.....[ny]' Law* | sed 's/:v.[ny]n.....[ny]//g' | sort | uniq -c > 01VerbInf.txt

egrep -o 'v.y.....[ny]' Law* | sed 's/:v.y.....[ny]//g' | sort | uniq -c > 01VerbRefl.txt

egrep -o 'v.[ny].....a[ny]' Law* | sed 's/:v.[ny].....a[ny]//g' | sort | uniq -c > 01VerbActive.txt

egrep -o 'nc[mf0]...' Law* | sed 's/:nc[mf0]...//g' | sort | uniq -c > 02NounCommon.txt

egrep -o 'nc[mf0].g.' Law* | sed 's/:nc[mf0].g.//g' | sort | uniq -c > 02NounGenitive.txt

egrep -o 'nc[mf0].d.' Law* | sed 's/:nc[mf0].d.//g' | sort | uniq -c > 02NounDative.txt

egrep -o 'nc[mf0].a.' Law* | sed 's/:nc[mf0].a.//g' | sort | uniq -c > 02NounAccusative.txt

egrep -o 'nc[mf0].l.' Law* | sed 's/:nc[mf0].l.//g' | sort | uniq -c > 02NounLocative.txt

egrep -o 'a[fr][mf][sp]...' Law* | sed 's/:a[fr][mf][sp]...//g' | sort | uniq -c > 03Adjective.txt

egrep -o 'r[rpcs][qmptc]' Law* | sed 's/:r[rpcs0][qmptc]//g' | sort | uniq -c > 04Adverb.txt

Extracting nominalizations *-šana* from the list of lemmas.

egrep -o '\b.*šana\b' Law* | sed 's/:n.....//g' | awk '{print $1 " " $2}' | sort | uniq -c > 05NMZ-šana.txt; the file was checked in nano to find and delete errors.

Rarely used grammatical forms (< 1000 per file) were excluded from the further analysis.

Checking the files in nano. Three files were edited in nano deleting the remained tags.

Merging files.

paste 01PartD.txt 01PartND.txt 01VerbActive.txt 01VerbBe.txt 01VerbCond.txt 01VerbDebitive.txt 01VerbInf.txt 01VerbMainInd.txt 01VerbRefl.txt 01VerbTikt.txt 02NounAccusative.txt 02NounCommon.txt 02NounDative.txt 02NounGenitive.txt 02NounLocative.txt 03Adjective.txt 04Adverb.txt 05NMZ-šana.txt -d "" > Laws.txt

Deleting the names of the source files.

sed 's/Law...//g' Laws.txt | sed 's/ b//g' > Laws2.txt

Adding names to columns.

nano Laws2.txt

Exporting file.

sftp > get Laws2.txt

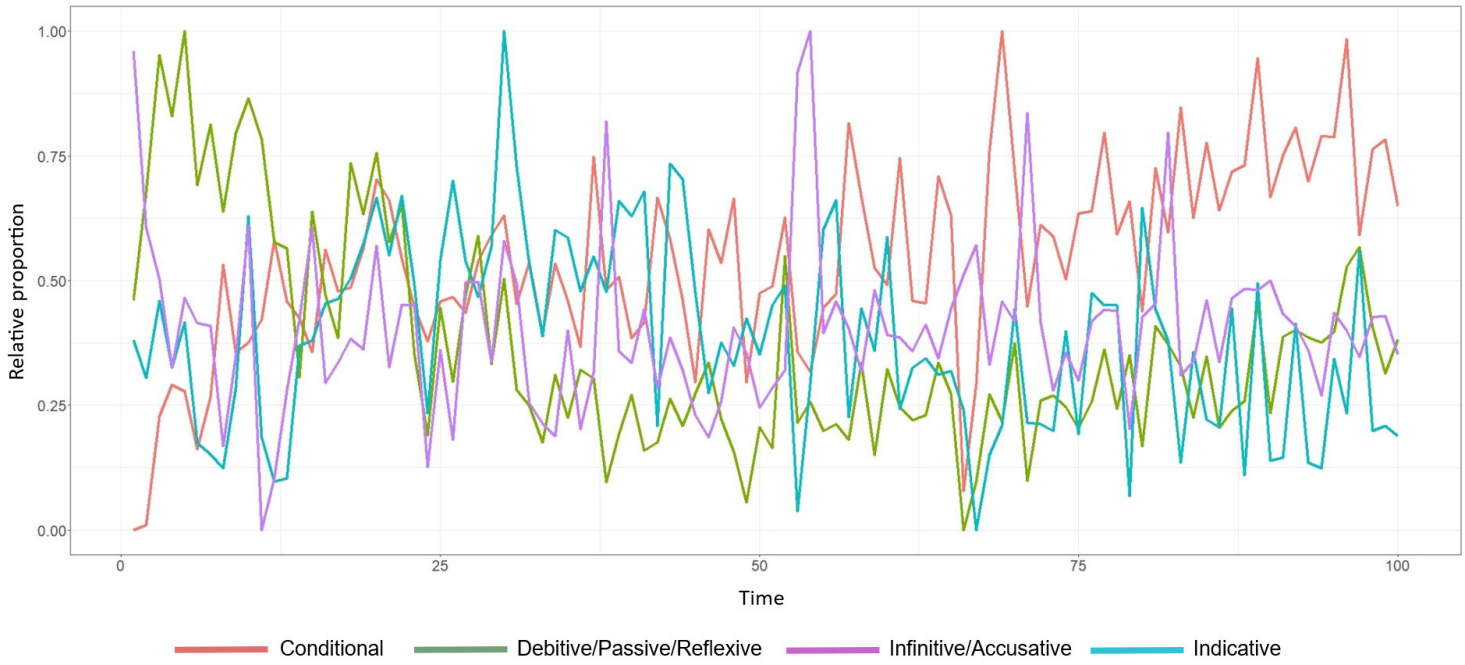

Figure 1. Diachronic dynamic of four scales after normalization, 1990–2022.

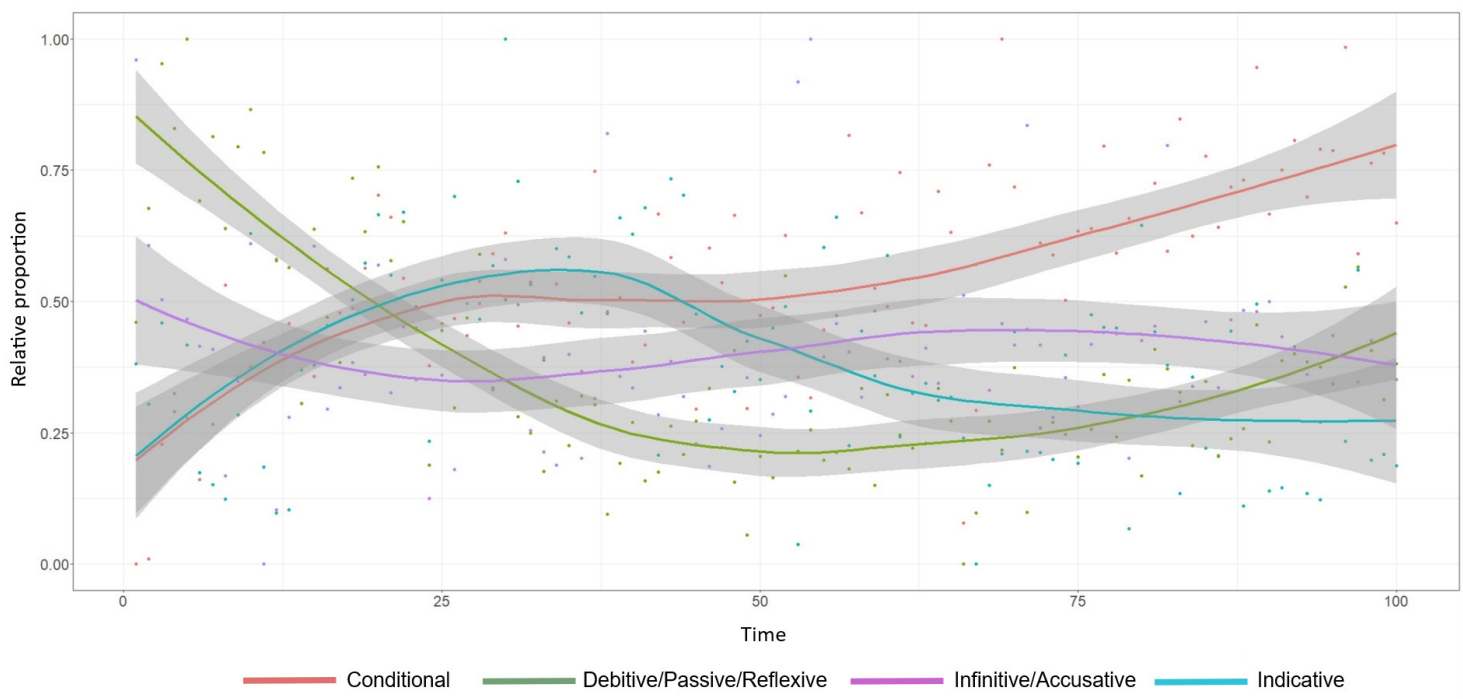

Figure 2. Locally estimated scatterplot smoothing of four scales after normalization, 1990–2022.