SEMANTIC ANNOTATION APPLIED FOR SUBJECTIVITY ANALYSIS OF ENGLISH-LANGUAGE ARTICLES FROM LATVIA'S CULTURAL CANON

KRISTĪNA KORNELIUSA

University of Latvia

Abstract. In cultural studies, subjectivity refers to self-recognition and manifestation of cultural identity, while in language, its lexical expression is directly related to the meaning of words denoting subjectivity. This study aims to explore the semantic fields relevant for subjectivity detection in texts on Latvian culture, using the semantic annotation method. A list of UCREL Semantic Analysis System (USAS) tags relevant for subjectivity detection was compiled based on subjectivity components and markers, and the differences in subjective tags across different thematic categories of texts of Latvia's Cultural Canon were explored. It was found that the most frequent subjectivity tags are on the positive side of the semantic scale, showing the significance of the included items; articles on film and visual arts refer to ideological aspects; expression of positive judgement of appearance is linked to the country's image; referring to subjective content not related to author's personal perspective is also a means of expressing subjectivity.

Key words: subjectivity, semantic annotation, Latvia's Cultural Canon, UCREL Semantic Analysis System, Wmatrix

INTRODUCTION

In linguistics, subjectivity refers to a verbal self-expression (Fina, 2009: 171), in cultural studies – to self-recognition of a cultural group (Sheikh, 2017: n. p.). Hence, the choice of words in texts that belong to the greatest examples of national culture is expected to convey subjectivity.

Subjectivity of a word may be defined by its meaning. The goal of the current study is to explore the semantic fields relevant to detection of subjectivity in English-language informative articles on Latvian culture. The following research questions are asked: (1) which semantic fields defined by the UCREL Semantic Analysis System (USAS) are relevant for subjectivity detection, and (2) how do the subjective semantic fields differ across the thematic categories of Latvia's Cultural Canon?

Semantic annotation is selected as a corpus analysis method with Wmatrix 5 as a research tool. The corpus for analysis consists of informative articles on

Latvia's Cultural Canon elements, which contains 108 texts in English from eight thematic categories and amounts to 70 109 words.

THEORETICAL BACKGROUND

1 SUBJECTIVITY DEFINITION

The meaning of subjectivity varies across 'different philosophical traditions and cultures' (Fina, 2009: 119), making it difficult to define from linguists' point of view. For instance, while House refers to subjectivity in language as 'the ability of a speaker to present himself or herself in and through language as a 'subject' and to indicate his or her attitude towards, and assessment of,' what he or she is saying (2012: 139), *subject* is a philosophical concept. Solomon describes subjectivity as a feature 'pertaining to the subject and his or her particular perspective, feelings, beliefs, and desires' (Solomon, 2005: n. p.). The clarification of what *subject* means is given by Peacocke (2012: 90), who states that being a subject refers to the capability 'of being in conscious states and of being the subject of conscious events'. Based on this claim, subject is a conscious being, and *subjectivity* – the experience of being one.

Subjectivity definition for corpus analysis calls for specific linguistic features. Antici et al. admit that 'spotting subjectivity is a challenging task, even for human experts' (2024: 273), implying that it is even more difficult when using digital tools. The scholars also point out that the *complexity and ambiguity* of perceiving subjectivity arises 'from different interpretations of the language, background knowledge, and personal biases' (ibid.). These biases are not characteristic of a corpus analysis tool, but can influence its functionality and/or the queries selected for detecting subjectivity.

Peng (2024: 8) states that 'in discourse, subjectivity bears five connotations', which are provided in Table 1.

Table 1. **Connotations of subjectivity in discourse constructions** (based on Peng, 2024: 8)

No.	Name	Explanation
1	'Communicative intention'	'wish, praise, blame, suggestion, and celebration'
2	'"Self" components of the speaker'	'perspective, emotion, opinion' etc.
3	'The speaker's stance and ideology'	(no additional explanation provided by Peng, 2024)
4	'The speaker's embodied experience and [context] perception'	'cognitive context, situational context, social context, cultural context' etc.
5	'Self elements [of the audience]'	'[The] 'cognitive states, social identities and social positions' [of the audience].'

Peng's connotations are in line with the definitions provided by other linguists and philosophers, who list the expression of attitude and assessment (House, 2012: 129) and 'evaluations and speculations' (Wiebe, Bruce, Martin, Wilson and Bell, 2004: 277) – the sets proposed by both linguists can be compared to such communicative intentions (see Table 1) as praise, blame and celebration; 'feelings, beliefs and desires' (Solomon, 2005: n. p.) can be linked to emotion, ideology and wish respectively. Lastly, self-components of a speaker include opinion, mentioned by Wiebe et al. (2004) and Baumgarten, Du Bois and House (2012).

There are researchers who view subjectivity as the first-person standpoint (Solomon, 2005), also called authorial stance (Pho, 2012), or, in Peng's case, speaker's stance. Subjectivity as authorial stance and its expression through the first-person singular pronoun has been explored using corpus analysis (e.g. by Korneliusa and Vinčela, 2023) and is, hence, omitted in the current research. However, ideology component of the third connotation is preserved for subjectivity analysis, since it is seen as the expression of personal beliefs.

It was decided to select Peng's first two connotations of subjectivity along with ideology for analysis, as the last two connotations refer to extratextual features. Based on terminology used by Peng in naming the second connotation – *self components of the speaker, subjectivity component* is a term adopted for all concepts listed as referring to subjectivity, as the paper focuses on the speaker/author-related connotations of subjectivity. It is taken into account that, although Peng (2024: 5-8) explicitly states that the selected methodology is corpus-based discourse analysis, the provided results are qualitative. To identify semantic tags denoting subjectivity across 21 semantic categories and obtain quantitative results, a more detailed list of subjectivity components has been created, namely: opinion, judgement, beliefs, attitude, evaluation and assessment, wishes and desires, feelings and emotions, based on Peng's explanation of connotations of subjectivity and subjectivity definitions by other scholars. The connotations of subjectivity may overlap; however, some dictionary definitions use them as synonyms for each other. For instance, *Cambridge Dictionary* (n.d.) defines *attitude* as 'a feeling or opinion about something or someone'. The process of attributing the semantic tags to each of these components is described in the section Procedure.

2 SUBJECTIVITY MARKERS

Subjectivity in language is conveyed through linguistic subjectivity markers. Peng (2024: 3) divides them into two groups: lexical – 'verbs, nouns, conjunctions, adjectives, and adverbs' and grammatical – 'modality, tenses, deixis, and [syntactic] constructions'. The grammatical markers except modality are excluded from this paper due to selecting semantic and not grammatical corpus

annotation. Modality, however, can be expressed with lexical means. For a more precise selection of subjectivity markers, the definitions of modality functions need to be listed (Love and Curry, 2021: 49) – 'hypothecality, possibility, instruction, intention, obligation, offering, permission, preference, promising, refusal, request, suggestion, ability, habit'

The discussion of subjectivity markers confirms the importance of modality in subjectivity analysis. Baumgarten et al. (2012: 2; 6) list personal pronouns, modal verbs and modal modifiers, specifying that modal adverbs include intensifiers; Pho (2012: 103-109) mentions personal pronouns, transitive verbs, the use of the determiner *this*, modal verbs and auxiliaries, 'attitudinal and epistemic stance adjectives, adverbs and nouns' and that-clauses; Fina believes modal adverbs and evaluative adverbs (2009: 171) are 'particularly relevant in [analysing subjectivity]' (ibid.: 134).

Overall, lexical markers of modality, as well as intensifiers, are added to the list of subjectivity markers.

3 THE RELEVANCE OF EXPLORING SUBJECTIVITY IN LATVIA'S CULTURAL CANON

Similarly to its philosophical and cognitive definition, subjectivity in cultural studies is seen as 'the condition of our being which enables us to recognize ourselves as subjects or persons' (Sheikh, 2017: n. p.). However, it is different in a sense that 'in cultural studies, [it] is believed to be culturally constructed' (ibid.). Hence, identifying subjectivity in texts on national cultural heritage allows us to explore the self-recognition of a cultural group.

To investigate the subjectivity expressed by Latvians, the articles on items included in Latvia's Cultural Canon (hereinafter – the Canon) are selected. The compilers characterise the Canon as 'a gateway to Latvia's cultural world' (Online 1). It has an English-language version, allowing foreigners to become acquainted with the Latvian culture. Subjectivity is explicitly stated in its introduction: 'the Canon is like *a pair of spectacles_through which* to catch sight of Latvia and Latvia's treasures'; '*through our eyes*, this unique selection becomes an inheritance [...]' (ibid., italics added in the examples) – signalling the presence of a particular perspective in the texts, which makes them relevant for analysing subjectivity.

The main purpose of the Canon is to inform and educate the public about the Latvian culture (Online 5). Among the 11 goals listed on the Canon's website, some create specific expectations regarding subjectivity markers. For instance, '[highlighting] new values' and '[bringing] to light forgotten elements' (ibid.) raise the possibility of high modality markers and intensifiers being present. Another goal is to '[raise] the prestige of culture', thus, words expressing evaluation, assessment or judgement are likely to bear a positive connotation. Finally,

the goal to '[unite] Latvian society around common features' (ibid.) is expected to be achieved by appealing to positive emotions. The frequency of emotion markers is not expected to be high, but the context they are found in is expected to highlight themes around which the Latvian society is united.

METHODOLOGY

The research corpus was compiled according to the following criteria – 'authenticity, representativeness, balance, sampling and size' (Korneliusa and Vinčela, 2023: 50). Its annotation involves 'adding interpretative linguistic information to a corpus' (Leech, 2004: n. p.). For the purposes of corpus study, semantic annotation of texts is selected. In recent years, it has been used for analysing information of various fields and topics, from news discourse (Peng, 2024) to healthcare data (Tiwari, Ortiz-Rodriguez and Jabbar, 2022). Tiwari et al. describe semantic annotation as 'a process of assigning relevant information to the concepts and their relationships'. The use of the term *concept* implies that it deals with the *meaning* of a word; Leech (2004: n. p.) refers to 'the semantic category of words'. According to the official website of UCREL (University Centre for Computer Corpus Research on Language at Lancaster University), the goal of semantic annotation is to distinguish between 'the lexicographic senses of same word' (Online 2) and a semantic tag – a mark attributed to a specific word or phrase as a result of semantic annotation – refers to words from the same *sense field* (ibid.), which corresponds to Leech's *semantic category*. The term *semantic field* is adopted in this paper (hereinafter referred to as SF).

PROCEDURE

1   CORPUS COMPILATION

The feasibility of choosing the Canon for subjectivity analysis has been briefly discussed above. Its representativeness in the cultural domain appears to be self-explanatory; however, it is important to mention that the Canon is supervised by the National Library of Latvia and supported by the Ministry of Culture of the Republic of Latvia (Online 1). Both its Latvian and English versions have been written by multiple authors. Most of the English-language texts are original, with just a few texts translated from Latvian. All the articles are seen as original, as the goal is not to analyse the quality of the translations but view all the texts as a way for Latvians to manifest their subjectivity internationally. Hence, the Canon can be perceived as an authentic and representative source for corpus compilation.

The Canon is divided into eight thematic categories – Traditional Culture, Architecture and Design, Performing Arts, Visual Arts, Literature, Music, Film,

and Landscapes. which correspond to the division of the corpus into eight subcorpora. No additional sampling, i.e., eliminating texts for the sake of balance was performed, as the research material needs to represent the entire Canon. The relative sizes of the subcorpora are presented in Figure 1. The final corpus consists of 70 109 words (the term *words* is used in the interface of the selected corpus analysis tool – Wmatrix5) across 108 texts.

Table 2. **Balance and structure of the research corpus**

Thematic category	Word count percentage
Traditional culture	9%
Architecture and design	15%
Performing arts	18%
Visual arts	12%
Literature	13%
Music	7%
Film	9%
Landscapes	13%

The corpus is not balanced: e. g., *Performing arts* is nearly twice as large as *Film*. The number of texts in each category is also unequal, ranging from 8 items in *Landscapes* to 18 items in *Architecture and design.*

As the texts are taken from a webpage, no optical character recognition-related issues had to be solved. The text from each article was copied to a Txt file, excluding the title and author/translator. The images and links were removed automatically upon copying the text to the Txt file.

2   UCREL SEMANTIC ANALYSIS SYSTEM AND WMATRIX TOOL

The UCREL semantic analysis system has been developed starting from the 1990s; a reference manual for all the assigned semantic tags was compiled in 2002 by Archer et al. However, the English C tagger using the system – Wmatrix, is regularly updated: Wmatrix4 was closed down on 1 March 2023; the version used in the current paper, Wmatrix5, will be available until April 2025.

The top level of the semantic tags hierarchy consists of 21 top-level labels or general discourse fields (Archer, 2002: 1). To each of them, a capital Latin letter is attributed. The attribution is either partially based on the name of the general discourse field (e.g., A – *abstract*, B – *body*, etc.), or the position of the letter in the English alphabet (the last letter – Z – used for words with limited semantic properties), or is seemingly random (e.g., K for entertainment, sports and games) which can be seen in Figure 2 (Archer et al, 2002: 2).

A general and abstract terms	B the body and the individual	C arts and crafts	E emotion
F food and farming	G government and public	H architecture, housing and the home	I money and commerce in industry
K entertainment, sports and games	L life and living things	M movement, location, travel and transport	N numbers and measurement
O substances, materials, objects and equipment	P education	Q language and communication	S social actions, states and processes
T Time	W world and environment	X psychological actions, states and processes	Y science and technology
Z names and grammar			

Figure 1. **The top level of USAS semantic tags hierarchy** (extracted from Archer et al., 2002: 2)

The 21 general discourse fields are subdivided into smaller categories – subdivisions (ibid.). The components of a USAS tag for the SF *Particular/general; detail* (ibid.: 5) is exemplified in Figure 3.

$$A4.2+$$

A – general discourse field
4 – first subdivision
2 – finer subdivision
+ – position on the semantic scale

Figure 3. **The components of a USAS semantic tag** (based on Archer et al., 2002: 1)

Each tag has a detailed description, including the prototypical examples as shown in Figure 4.

A4.2   Particular/general; detail
General/abstract terms denoting (level of) generality/detail

PROTOTYPICAL EXAMPLES: CLEAR-CUT (+), DETAILED (+), PARTICULAR (+),
GENERALISE (-), GENERIC (-), INDEPT (+), VAGUE (-)

ATTENTION TO DETAIL (+), GET TO THE POINT (+),
IN THE BROADEST TERMS (-), IN GENERAL (-),
ON A BROAD SCALE (-), THE INS AND OUTS (+)

Figure 4. **An entry describing a USAS semantic tag** (extracted from Archer et al., 2002: 5)

Looking at the prototypical SF, it is also possible to see that ' + ' and '–' do not necessarily refer to positivity/negativity, but the position on the semantic scale. An illustration for a semantic scale has been created and is presented in Figure 5.


-
Vague   Neutral   +
                          Particular/Detailed
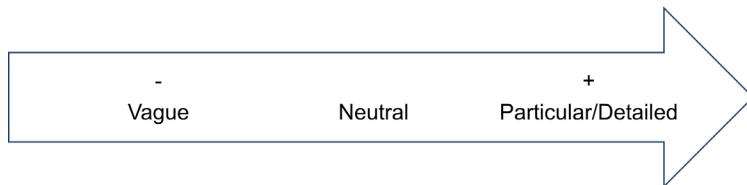
Figure 5. **A USAS semantic scale for the tag A4.2**

The words on the left-hand side of the spectrum, i.e., marked with A4.2– will refer to vagueness, e.g., *generalise, vague, in the broadest terms* (Archer, 2002: 5). Note that several minuses can also be added, depending on the extent of vagueness. The words on the left-hand side of the spectrum, i.e., marked with A4.2 + (or with several pluses) will refer to a certain extent of detail, e.g. *clear-cut, particular, attention to detail* (ibid.).

3   THE SELECTION OF SUBJECTIVITY MARKERS FROM SEMANTIC TAGS

The semantic tags were selected based on the following criteria derived from theory:
1) the relevance of the name and general description of the SF to the subjectivity components – opinion, judgement, beliefs, attitude, evaluation/assessment, wishes/desires, feelings/emotions;
2) the relevance of the name and general description of the SF to the subjectivity markers;
3) the relevance of the SF to the subjectivity expression in the selected context.

The reasoning for the selection based on subjectivity components is presented in Table 2. The tags are arranged according to the subjectivity components. Additional comments for explaining the attribution are provided.

Table 3. **The selection of semantic tags relevant for subjectivity detection based on the definition of subjectivity**

Expression of subjectivity	Tag	Name	Comment on the SF(s)
Attitude	A1.1.2	Damaging and destroying	bears a negative connotation
	A1.9	Avoiding	
	S1.1.2	Reciprocity	refers to attitude displayed in a relationship between the subjects
	S7.2	Respect	
	S8	Helping/hindering	
Assessment/ Evaluation	A5 (including all the sub-categories)	Evaluation	explicitly stated in the name
	A1.2	Suitability	the extent of both is determined according to subjective criteria
	A1.5.2	Usefulness	
	G2.1 + /−	Crime, law and order: Law & order	assessment/evaluation is performed according to specific criteria – criminal code; only the tags marked on a semantic scale are included
	S1.1.4	Deserve etc.	the entitlement/eligibility/ merit are determined according to specific criteria
Emotions/ Feelings	E (including all the sub-categories)	Emotional actions, states & processes	explicitly stated in the name
	X5.2	Interest/boredom/excited/ energetic	the name refers to emotions/ feelings
Judgement	O4.2	Judgement of appearance (pretty etc.)	explicitly stated in the name
Belief	X2.1	Thought, belief	
	G2.2 + /−	General ethics	the name refers to a set of beliefs; only the tags marked on a semantic scale are included.

Expression of subjectivity	Tag	Name	Comment on the SF(s)
Wish/desire	X7	Wanting; planning; choosing	the name refers to wishes/ desires
	X2.6	Expect	the name and description refer to wishful thinking
Opinion	A8	Seem/Appear	the name and description refer to opinion

As some tags could not be included in the table above, due to the fact that the subjectivity connotations they refer to overlap (see the discussion of *Cambridge Dictionary* (n.d.) definition in Section 1), they are presented in Table 3.

Table 4. **The selected tags which can be attributed to subjectivity components**

Tag	Name	The subjectivity component and/or additional comment (if relevant)
A1.7	Constraint	*Attitude, assessment, opinion.* The predominance of words marked on the semantic scale in the prototypical SF. The tag A11 is included with all the subcategories.
A6.2	Comparing: Usual/Unusual	
A11	Importance	
A12	Easy/Difficult	
B2 + /–	Health and disease	*Attitude, feeling.* Only words marked with a + or – are relevant.
S1.2	Personality traits	*Attitude, opinion, judgement.* Included with all the subcategories.
X2.5	Understand	*Attitude, opinion*
X4.1	Mental object: Conceptual object	*Attitude, opinion, emotion, evaluation/ assessment.* The words from the prototypical SF repeat the words from subjectivity definitions explored.

Both Table 2 and Table 3 show the selection of subjectivity markers based on the subjectivity components identified in Section 1. Meanwhile, the names and descriptions of some semantic tags directly refer to modality or modality functions:
- A7: Definite (+ modals)
- S6: Obligation and necessity;
- S7.4: Permission;
- X9: Ability

They have been included as relevant for subjectivity analysis.

Another set of relevant semantic tags is A13 and its subcategories, and A14 and its subcategories. A13 – *Degree* – includes intensifiers and other related discourse markers. This can be linked to the discussion by Baumgarten et al. (2012: 2). A14 contains *Exclusivizers/particularizers*, and the function of the words in this SF is 'to draw attention to/focus upon X' (Archer et al., 2022: 9). It is assumed that drawing attention to something is the subject's personal choice.

4　DATA EXTRACTION AND VISUALISATION

Using Wmatrix5, the corpus files were arranged into eight folders according to the thematic category (see Figure 6).
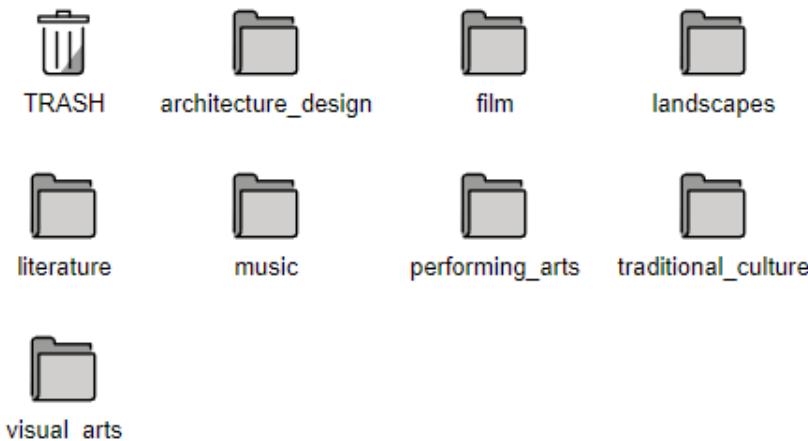
Figure 6. **Subcorpus folders.** Wmatrix5 interface

Due to the tool functionality, all texts from a category had to be copied in a single file. For this reason, when providing an example, it is not possible to refer to a specific text. All the examples quoted in the Discussion of the Results are referred to by the name of the category.

Next, for each folder, two files were extracted – *Semantic frequency list,* showing the frequency of all the semantic tags, and *Word and USAS frequency list,* providing the word frequencies and semantic tags (see Figures 7 and 8).

Kristīna Korneliusa. SEMANTIC ANNOTATION APPLIED FOR SUBJECTIVITY ANALYSIS ..

```
TOTAL               10366
Z5                   3476
Z99                   584
Z8                    447
N1                    296
Z2                    254
H1                    251
C1                    250
A3+                   219
M6                    124
M7                    124
A1.1.1                116
Z1                    115
T1.3                  109
A9+                    81
H2                     72
Z3                     71
N5++                   68
O2                     67
T1.1.2                 67
O1.1                   67
P1                     65
A2.1+                  62
A13.3                  58
A11.1+                 56
M1                     54
```

Figure 7. **Semantic frequency list for Architecture and design subcorpus**

```
TOTAL               10366
the                   Z5            896
and                   Z5            401
of                    Z5            392
in                    Z5            276
a                     Z5            232
to                    Z5            182
for                   Z5            113
is                    A3+            86
was                   Z5             84
latvian               Z2             66
with                  Z5             66
it                    Z8             66
by                    Z5             66
was                   A3+            65
as                    Z5             59
on                    Z5             57
from                  Z5             57
buildings             H1             55
he                    Z8             54
which                 Z8             53
its                   Z8             52
his                   Z8             52
an                    Z5             49
at                    Z5             49
Rga                   Z99            41
also                  N5++           40
```

Figure 8. **Word and USAS frequency list for Architecture and design subcorpus**

The lists were copied into MS Excel workbooks and the frequency column was manually separated from the tag and/or word column. The RF was calculated per 10 000 words, considering the sizes of the subcorpora, ranging from 5 236 words in *Music* to 12 273 in *Performing arts*. The specific words from subjective semantic fields (based on the frequency results; see Results) were arranged in *Word and USAS frequency list* Excel workbook. If the content of the semantic field provided an insight to be discussed, it was visualised using TagCrowd – a list of words, considering their absolute frequency, was entered as a query (see Figure 9).
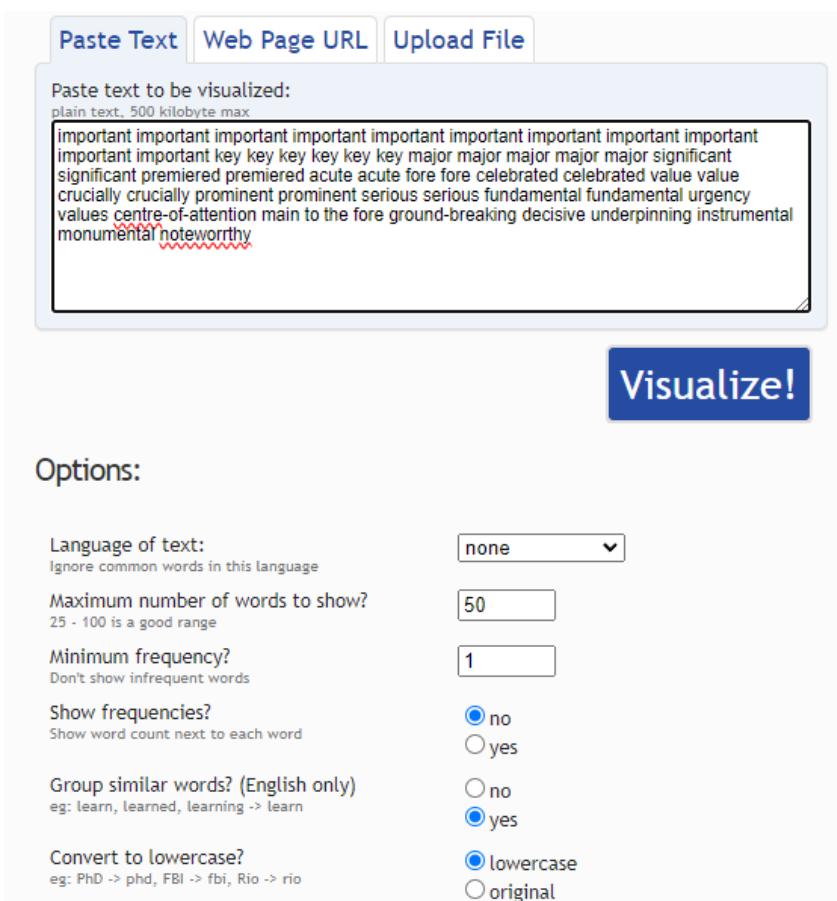
Figure 9. **TagCrowd query for the SF A11.1+ in the category** *Performing arts*

Multi-word units, such as *centre of attention,* had to be hyphenated to be counted as a single unit. Absolute frequencies were omitted from the final image, because what matters is the relative size of the words for illustration.

The extraction of relevant contextual information for the USAS tags was performed using *Concordance – USAS tag* function in Wmatrix5. The query using a specific tag was entered, and the concordance lines were extracted (Figure 10).

Figure 10. **Concordance of the tag O4.2+ in the category *Landscapes***

DISCUSSION OF RESULTS

The three most frequent tags across the thematic categories are shown in Table 4.

Table 5. **The three most frequent tags across the thematic categories of the Canon** (per 10 000 words)

Category	Tag 1	RF	Tag 2	RF	Tag 3	RF
Traditional culture	E4.1+	62	A7+	60	A6.2+	56
Architecture and design	A13.3	56	A11.1+	54	A7+	47
Performing arts	A7+	49	A13.3	46	A11.1+	44
Visual arts	X4.1	54	A11.1+	48	X2.1	45
Literature	A13.3	53	A7+	49	X2.1	40
Music	A13.3	36	A11.1+	34	E4.1+	32
Film	A13.3	62	A7+	50	X2.1	41
Landscapes	A11.1+	78	O4.2+	75	A13.2	53

The words from the SFs listed in Table 5 were explored and visualised across all the categories. Additionally, the second most frequent tag denoting

emotions E2+, found 33 times per 10 000 words in *Literature*, 21 times in *Music* and 21 times in *Performing arts*, was explored.

The three semantic tags most often found across all the categories are A7+, A13.3, and A11.1+. These SFs are constituted from largely the same words across all the categories – see their visualisation in Appendix 1. Boosters – 'intensifiers [amplifying] to a high degree' (Archer et al, 2002: 8) – were more frequent than the other types of degree markers from the field A13, and both the modality markers and terms denoting importance or significance are found on the right-hand side of the semantic scale. This is in line with the expectations based on the Canon goals.

It was expected that the SF O4.2 (judgement of appearance) would be found more frequently in categories with a prominent visual component – *Architecture and design*, *Visual arts*, *Film* and *Landscapes*. The data, however, suggest that O4.2+ is the second most frequent one in *Landscapes*; in *Architecture and design* and *Visual arts* it is only slightly more frequent than in *Music*, while in *Film*, it is not even among the 10 most frequent subjective SFs. This is due to the fact that in the articles on the Latvian film industry and visual arts, thoughts, beliefs and ideology are highlighted more prominently than the visual aspects (SF X2.1) – consider the authors' personal views on the film or director discussed (see Example 1 (E1) and other examples referred to in Appendix 4), the historical view of the Latvian filmmakers from the ideological point of view (E2) and the film characters' thoughts in a plot description (E3). Likewise, the articles on visual arts emphasise general views about a painting (E4) or opinions and impressions of art historians (E5 and E6).

The SF O4.2+ in *Landscapes* is illustrated in Figure 11.

Figure 11. **The SF Judgement of appearance in the category *Landscapes***

The relatively high frequency of the words *impressive*, *beautiful*, *beauty*, *pic-turesque* is in line with the summary by the Canon compilers, stating that 'the eight canonical landscapes represent Latvia's beauty and diversity' (Online 4). Thus, the texts aim not only to describe the visual aspects of nature but provide a visual representation of the country – see E7 and E8. The adjectives marked as O4.2+ modify nouns – *perspectives*, *landscapes* – contextually related to specific place names – *Kandava*, *Sabile*, *Kurzeme*, *Daugava* – which further explains why the visual description of the natural scenery is bound to the country's image.

The categories *Traditional culture* and *Music* contain more words related to emotions than the other categories. In both cases, in the SF E4.1 +, the word *celebration* in both its singular and plural form is the most frequent (see Appendix 2). Since in this paper, words denoting emotions are seen as uniting the nation around common values, it can be concluded that common celebrations are one of such values. Moreover, the word is found in the same SF *joy* and *happy*, which refer to positive emotional states.

The second most frequent marker of emotions, which is likewise positive – E2 + (*Liking*) – is found in the categories *Literature, Music* and *Performing arts*.

There is a larger variety of words from E2 + in *Literature* than *Music* (see Appendix 4). The E2 + words in *Literature* can be seemingly attributed to SF ROMANCE {love, fall for, best-loved, loving, beloved, amorous, lust, devotion}. The context, however, leads to the following findings (see the examples referred to in Appendix 4):

- *Devotion* and *best-loved* is used outside the romantic context (E9 and E10).
- *Love*, *fall for* and *beloved* are used to retell the plot of a novel where romance is involved (E11 from the article about the novel *Aija*).
- *Amorous* is used as a modifier to describe the nature of the events in a novel (E12, referring to *Aija*).

This suggests that subjectivity of a text is influenced not only by the author's emotions, but also the content referred to, e.g., novels about an emotionally loaded relationship.

The semantic field E2 + in *Performing arts* includes words *liked*, *loved*, and *adored* (see Appendix 3), which are arguably of different intensity; however, their position on the semantic scale is the same. Further research is necessary to claim that a more detailed lexicon is necessary to obtain more precise results regarding the level of subjectivity. However, this is out of the scope of the current paper.

It is arguable whether a reference to someone else's opinion (e.g. art historians referred to in *Visual arts*) or a content of a novel (*Aija* in *Literature*) are to be seen as expression of authors' subjectivity. The author of the current research sees these cases as a proper expression of subjectivity which should not be separated from the personal views of the Canon authors – referring to subjective content is also a means of expressing it.

CONCLUSIONS

The paper has explored the manifestation of cultural subjectivity in language using semantic annotation of texts from Latvia's Cultural Canon. Based on theoretical considerations, a list of USAS tags denoting subjective words was created. The connotations of subjectivity defined by Peng (2024) were taken into account and linked to more specific subjectivity components – opinion, judgement,

beliefs, attitude, evaluation/assessment, wishes/desires, and feelings/emotions. Occasionally, the meanings of these components intersect. Intensifiers and lexical markers of modality were added to the list of subjectivity markers.

The most frequent SFs across the eight thematic categories of the Canon are abstract terms of modality, abstract terms denoting importance or significance, and intensifiers amplifying to a high degree. These tags, along with markers of emotions and judgement of appearance, were found on the positive side of the semantic scale. The results are in line with the goal of the Canon to highlight the significance of the included items and unite the nation around common values.

The analysis of concordance lines provides valuable insights regarding the differences in subjective SFs across categories. Judgement of appearance is most frequently expressed in *Landscapes*, as the visual aspect of nature is crucial for the country image. In *Visual arts* and *Film*, the ideological component is more prominent. The markers of emotions in *Literature* show that subjectivity of a text is influenced not only by the author's personality, but also the choice of the subject matter. The author of this paper reckons that referring to subjective content is likewise a means of expressing subjectivity.

Finally, further research is suggested with a view to create a more detailed lexicon, considering the intensity of words currently marked by the same semantic tag. The adjustments will yield conclusions on the level of subjectivity.

ACKNOWLEDGEMENTS

REFERENCES

Antici, F., Galassi, A., Ruggeri, F., Korre, K., Muti, A., Bardi, A., Fedotova, A. and Barrón-Cedeño, A. (2024) A Corpus for Sentence-Level Subjectivity Detection on English News Articles. *LREC-COLING 2024*: 273–285. Available from https://aclanthology.org/2024.lrec-main.25.pdf [Accessed on 24 May 2024].

Archer, D., Wilson A., and Rayson, P. (2002) Introduction to the USAS Category System. Available from https://ucrel.lancs.ac.uk/usas/ [Accessed on 2 June 2024].

Baumgarten, N., Du Bois, I. and House, J. (2012) Introduction. In N. Baumgarten, I. du Bois and J. House (eds.) *Subjectivity in Language and in Discourse* (pp. 1-14). Bingley: Emerald Group Publishing Limited.

Fina, A. (2009) Language and Subjectivity. *Estudios de Lingüística Aplicada*, 27 (50): 117-176.

House, J. (2012) Subjectivity in English lingua franca interactions. In N. Baumgarten, I. du Bois, J. House (eds.) *Subjectivity in Language and in Discourse* (pp. 139-156). Bingley: Emerald Group Publishing Limited.

Korneliusa, K. and Vinčela, Z. (2023) The Application of Diachronic Corpus Compilation Principles in a Pilot Study of Subjectivity. *Baltic Journal of English Language, Literature and Culture*, 13: 48-68.

Leech, G. (2004) Adding Linguistic Annotation. In M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice.* Available from https://books.google.lv/books/about/Developing_Linguistic_Corpora.html?id=41ZJAAAAYAAJ&redir_esc=y [Accessed on 25 May 2024].

Love, R. and Curry, N. (2021) Recent Change in Modality in Informal Spoken British English: 1990s-2010s. *English Language & Linguistics*, 25 (3): 527-562. doi:10.1017/S1360674321000265.

Peacocke, C. (2012) Subjects and Consciousness. In A. Coliva (ed.) *The Self and Self-Knowledge* (pp. 74-101). Oxford: Oxford University Press.

Peng, B. (2024) Subjectivity of Discourse Constructions in News Discourse by Integrating Construction Grammar and Critical Discourse Analysis. *Applied Mathematics and Nonlinear Sciences*, 9 (1): 1-16. doi:10.2478/amns-2024-1050.

Pho, P. D. (2012) Authorial Stance in Research Article Abstracts and Introductions from Two Disciplines. In N. Baumgarten, I. du Bois and J. House (eds.) *Subjectivity in Language and in Discourse* (pp. 97-114). Bingley: Emerald Group Publishing Limited.

Sheikh, F. A. (2017) Subjectivity, Desire and Theory: Reading Lacan. *Cogent Arts & Humanities*, 4 (1299565). doi:0.1080/23311983.2017.1299565.

Solomon, R. (2005) Subjectivity. In T. Honderich (ed.) *The Oxford Companion to Philosophy* (p. 900). Oxford: Oxford University Press.

Tiwari, S., Ortiz-Rodriguez, F. and Jabbar, M. A. (2022) Semantic Modeling for Healthcare Applications: An Introduction. In S.Tiwari, F. Ortiz-Rodriguez and M. A. Jabbar (eds.) *Semantic Models in IoT and eHealth Applications* (pp. 1-17). doi:10.1016/B978-0-32-391773-5.00007-8.

Wiebe, J., Bruce, R., Martin, M., Wilson, T. and Bell, M. (2004) Learning Subjective Language. *Computational Linguistics*, 30 (3): 277-308.

DICTIONARIES

Cambridge Dictionary (n.d.). Available from https://dictionary.cambridge.org/dictionary/ [Accessed on 5 June 2024].

INTERNET SOURCES

[Online 1] Available from https://kulturaskanons.lv/en/par-kanonu-2/ [Accessed on 25 May 2024].

[Online 2] Available from https://ucrel.lancs.ac.uk/annotation.html [Accessed on 25 May 2024].

[Online 3] Available from https://ucrel.lancs.ac.uk/wmatrix/ [Accessed on 2 June 2024].

[Online 4] Available from https://kulturaskanons.lv/en/list/ [Accessed on 25 May 2024].

[Online 5]Available from https://kulturaskanons.lv/en/merki/ [Accessed on 7 May 2024].

AMALYSED TEXTS

Available from https://kulturaskanons.lv/en/list/ [Accessed on 25 May 2024].

TOOLS

TagCrowd. Available from https://tagcrowd.com/ [Accessed on 9 June 2024].

Wmatrix5. Available from https://ucrel-wmatrix5.lancaster.ac.uk/wmatrix5.html [Accessed on 9 June 2024].

APPENDIX 1
THE VISUALISATION OF THE SFS IN THE CATEGORY *PERFORMING ARTS*

A7+

can cement cementing certainly clear coherent could ensure made-sure may potential would

A13.3

a-lot deeply extremely fiercely greatly heavily highly hugely indeed more much particularly really so strongly very

A11.1+

acute celebrated centre-of-attention crucially decisive fore fundamental ground-breaking important instrumental key main major monumental noteworthy premiered prominent serious significant to-the-fore underpinning urgency value values

APPENDIX 2
THE SF E4.1+ IN ARTICLE CATEGORIES TRADITIONAL CULTURE AND MUSIC

Traditional Culture

celebration celebrations

Music

celebrated celebration celebrations happy joy

APPENDIX 3
THE SF E2+

Literature

adored affinity amorous appreciated beloved best-loved cherished dear devotion enjoyed fall-for like lived-with loved loving lust popular popularity takes-to

Music

affections beloved cherished dear gloried loved **popular**

Performing Arts

adored applause appreciated darling enjoyed fan fondly **liked loved** penchant **popular** precious

APPENDIX 4
A LIST OF EXAMPLES OF WORDS IN CONTEXT

[1] 'It is incredible to *consider* that this was Piesis' first time directing a feature length fiction film.' (*Film*)

[2] 'Latvians were not *trusted* to make films from an ideological and technical perspective.' (*Film*)

[3] 'When all hope is *thought* to be lost an Estonian fishing boat sails past.' (*Literature*)

[4] 'Janis Rozentāls' painting After the Service [...] is widely *considered* the iconic Latvian painting.' (*Visual arts*)

[5] 'The art historian Vilnis Vējš [...] *finds* it significant that it is the gateway that is often the subject matter.' (*Visual arts*)

[6] 'The art historian Laima Slava [...] *believes* that the intention of the photographer was [...] to catch the *feeling* of living there.' (*Visual arts*)

[7] 'The *beautiful*, wide, and colourful perspectives served as a reason [...] for the river valleys section from Kandava to Sabile to be metaphorically called the Switzerland of Kurzeme.' (*Landscapes*)

[8] 'Here the *magnificent* landscapes of Daugava were formed by the steep gradient of the river with many rapids and restless flow, the *beautiful* cliff landscapes [...].' (*Landscapes*)

[9] '[Imants Ziedonis'] *devotion* to spirituality and dignity is continued by the Imants Ziedonis Fund [...]'. (*Literature*)

[10] 'If Rainis [...] is Latvia's most esteemed writer, Kārlis Skalbe would have a strong claim to be regarded as its *best-loved*.' (*Literature*)

[11] 'In Echo, Jānis' *beloved* Aija remains all but invisible; what makes him [...] *fall for* her again are memories of a *love* that even his young shepherd self had recognized to be misguided.' (*Literature*)

[12] 'Suddenly we see characters tackling class and economic matters in the manner of Hamsun [...] alongside things *amorous* and fatalistic.' (*Literature*)

Kristīna Korneliusa is currently a doctoral student working at the University of Latvia. Her research interests include corpus linguistics and discourse analysis.

iD https://orcid.org/0000-0001-5003-5445

Email: kristina.korneliusa@lu.lv