Computational recognition of narratives

Applying narratological definitions to the analysis of political language use

Mari Hatavara*, Kirsi Sandberg, Mykola Andrushchenko, Sari Hälikkö, Jyrki Nummenmaa, Timo Nummenmaa, Jaakko Peltonen and Matti Hyvärinen

Tampere University

*Tampere University and University of Tennessee

Abstract

Computational recognition of narratives, if successful, would find innumerable applications with large digitized datasets. Systematic identification of narratives in the text flow could significantly contribute to such pivotal questions as where, when, and how narratives are employed. This paper discusses an approach to extract narratives from two datasets, Finnish parliamentary records (1980–2021) and oral history interviews with former Finnish MPs (1988–2018). Our study was based on an iterative approach, proceeding from original expert readings to a rule-based, computational approach that was elaborated with the help of annotated samples and annotation scheme. Annotated samples and computationally found extracts were compared, and a good correspondence was found. In this paper, we exhibit and compare the results from annotation and rule-based approach, and discuss examples of correctly and incorrectly found narrative sections. We consider that all attempts at recognizing and extracting narratives are definition dependent, and feed back to narrative theory.

Keywords: narrative theory, recognition of narratives, computational approach, rule-based searches, annotation, digital humanism, parliamentary records, applied narratology

Introduction

In this article, we explore how narratological definitions and theories of narrative can be applied to recognizing narratives in naturally occurring political language use with computational methods. As a next step, the reliable recognition of narratives in large databases would enable targeted narratological analysis in various social and political contexts. Existing computational approaches to narratives aim at modeling narrative structures for different purposes, such as creating automated systems to generate and interpret narratives as well as supporting literary scholars with their analyses (see Mani, 2014, Bögel et al., 2015, Ek & Wiren, 2019; Miller & Park, 2020, Piper et al., 2021). Studies of detecting narratives from non-narrative and non-fictional data, particularly clinical texts (Yang & Garibaldi, 2015, Lin et al., 2020) and newspaper articles (Eisenberg & Finlayson, 2016, Sudhahar et al., 2011, Zhang et al., 2019), are often based on extracting chronological sequences of actions or events from the text, i.e. by mapping explicit "semantic triplets" of subject-action-objects. Existing research also recognizes the role played by subjective language and agency in data such as interviews and personal weblogs (Sagae et al., 2013; Andrade & Andersen, 2020). Work presented in this paper undertakes a novel objective to exploit grammatical features in a versatile manner in narrative detection as well as to cultivate the precision to distinguish narratives from other types of action portrayal, such as general reporting of past events. This approach is needed particularly in addressing language use in contexts where argumentative and persuasive functions surpass storytelling aims, such as parliamentary talk.

The basic theoretical problem in the recognition of narratives concerns the definitions of the concept, as there is no consensus on how to define narrative (Abbott, 2002, pp. 12–24, Tammi, 2006). This article brings together essential definitions of narrative in order to operationalize those for digital applications. While much of the previous research has concentrated on analyzing the story contents, our approach recognizes the act of telling as an equally crucial component of narrative: narratives are situated tellings of a story (Herman, 2002; 2009). Moreover, we recognize both main trends in the definitions of narrative, the one based on the existence of two events (Prince, 1982) and the other emphasizing the mediation of experience (Fludernik, 1996). Recognizing narratives as always entailing both the telling

and the told, as well as incorporating both event-centered and experience-centered narratives makes our approach stand out from other approaches aiming at computationally recognizing narratives. Furthermore, definitions of narrative often contain cognitive and discursive elements that cannot easily be reduced to sentence-level linguistics, such as "surprise" or "disruption in the storyworld" (Herman, 2009). Readers and annotators in our case easily recognize these interpretative qualities, but operationalizing them for automated approaches is challenging.

Computational recognition of narratives can proceed along two tracks, each with their inherent challenges. One of them is a machine learning process that operates without predefined rules, relying on a large corpus of materials manually annotated as narratives and non-narratives. An examination of them enables a neural network to separate one from the other and apply this separation to other corpora, too. The challenge with machine learning is that it requires a vast amount of annotated training data before it can accurately tell narratives apart from non-narratives. The other solution is a rule-based approach, where exact linguistic features are defined and operationalized for algorithms. The design of such features for rule-based algorithms is, in turn, complicated, since narrative cannot be reduced to sentence-level phenomena but requires a broader, discourse-level examination. Also, as Tuija Virtanen (1992) maintains, linguistic categories tend to be fuzzy and transform in use. Many narratologists argue that narrative may rather be a "more-or-less" than an "either-or" concept (Herman, 2009, pp. 12–17). To understand this gradience of narrative, narratologists employ the concept of "narrativity" to denote narrative features that may exist also in a text which as a whole is not identified as narrative (see Prince, 1987, p. 65).

Our solution is to combine the design of a rule-based algorithm with a moderate amount of expert annotation data. Linguistic features are deduced from narrative theory and are iteratively evaluated by and improved in comparison to the annotated data. Creating an annotation scheme that elaborates features sufficient for a narrative is far from easy in itself, as the annotation scheme entails unambiguous qualifications for interpreting a section as a narrative (cf. Reiter et al., 2019). Still, in the hands of expert annotators it helps to bridge the transfer from the definitions of narrative to the linguistic rules in the algorithm and its results enable curating the results of the algorithm. The development process is schematically

represented in Figure 1, where solid lines indicate initial "one-time" contributions and dashed lines refer to iterative, continuous interactions.

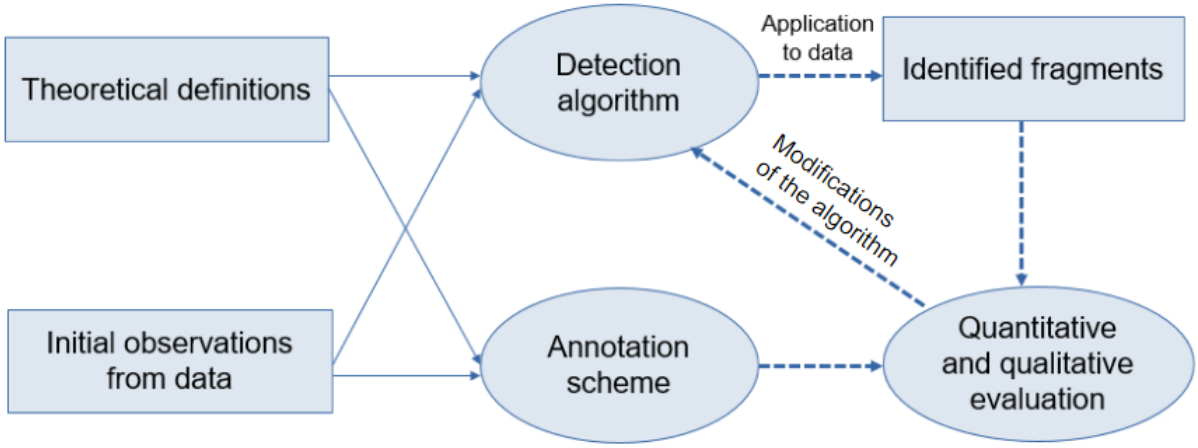

Figure 1. Development of a narrative-finding algorithm

Our multidisciplinary team works together on parliamentary data from two sources: Finnish parliamentary records (1980–2021) and transcribed oral history interviews of former Finnish MPs (1988–2018). These materials help to evaluate the applicability of our recognition algorithm in different communicative environments. Both datasets have been produced and processed by the Library of Parliament and Archive of Parliament in Finland, therefore following the same rules of transcription and avoiding inconsistencies in the results that may be caused by different processing methods. From first test runs with preliminary versions of the algorithm we learned that the model worked fairly well with the interview data but less satisfactorily with the parliamentary records.

This article proceeds by describing our data as well as the search tool enabling the use of algorithms for analysis. Next, we present the narrative-theoretical basis for our approach as well as our procedure with the annotation and the algorithm. After that, we present the quantitative results, including the number of narratives found and the correspondence between annotators and the algorithm. We then discuss critical cases about narratives identified by both annotators and the algorithm, as well as narratives and non-narratives on which they disagree. In conclusion, we evaluate the results and discuss how (Finnish) language-dependent they are.

Our database of parsed data and a search tool

The data of the present study is drawn from two datasets of transcribed political talk: Finnish parliamentary records (1980–2021; 93 million word tokens) and oral history interviews with former Finnish MPs (1988–2018; 12 million word tokens). Parliamentary talk is highly regulated by procedural rules and also argumentative in nature, weighing the consequences of alternative choices. The interviews, in turn, address the life course and particularly the political career of former MPs, portraying past events often with hindsight. In the long, semi-structured interview conversations former MPs cover a fixed selection of topics centered around their personal and political biography, networks and personal views. Both the interviews and the parliamentary records are transcribed from spoken mode to standard written Finnish by the Library of Parliament. Due to this procedure, certain features of spontaneous speech (such as self-corrections) and morpho-phonological variation of regional dialects are systematically changed, while regional and social variation in word choices and personal style is not. Guidelines of this process seek authenticity, consistency, and reliable account of conversations (Voutilainen, 2017).

We have organized parliamentary records and interviews into a database (VoDe Corpora, 2021). Each entry relates a single sentence to metadata about the record's time, its position in the document, the speaker's identity and political party. In order to operationalize the grammatical structure of sentences, texts were parsed and annotated by the Finnish dependency parser (Haverinen et al., 2014). Each lemma is accompanied by part-of-speech tagging (noun, verb, adjective, etc.), its inflectional features (conjugation or declension) and syntactic function (subject, object, etc.). Moreover, we utilized the theoretical framework of lexical argument structures (see Rappaport Hovav & Levin, 2015) to separate and classify speech act verbs (300 lexemes) out of ca. 4500 verb lexemes from our interview data into a semantic resource for our model.

The search tool we have developed provides a possibility to conduct searches based on grammatical features of words and sentences as well as metadata. What is more, we improved the dependency parser's identification of verb tenses involving compound forms. Our unique data with the built-in search system allows us to pilot a model for computerized analysis of

texts based on the combination of lexical, syntactic and semantic features together with metadata. For a more detailed account on features of the database and the search tool see Andrushchenko et al. (2021).

Narrative definitions operationalized

Many definitions of narrative focus on the notion of event and its subsequent representation. An early example would be Gérard Genette's (1980): "any narrative - - is a linguistic production undertaking to tell of one or several events" or Gerald Prince's (1982): "[Narrative is] the representation of at least two real or fictive events or situations in a time sequence". Noël Carroll (2001, pp. 21–22 and in passim.) emphasizes the narrative connection as the most important characteristic of narrative. This narrative connection between at least two events or states of affairs is generated when the temporal relations between the events are clearly expressed, the earlier events contribute to the later ones, and the events concern at least one unified subject (p. 32). Carroll's (pp. 24–25) discussion of the distinction between narratives and chronicles – building on Hayden White's (1987) discussion of the terms – is also helpful in our annotation, where chronicles have their own code. For Carroll (pp. 25–26) the main distinguishing feature of a chronicle is that, in spite of it portraying events in temporal succession, no other meaningful connection between them is displayed as is the case in narratives. This connection can be pinpointed more specifically by recognizing that the mere existence of two events or "temporal juncture" between narrative clauses (Labov & Waletzky, 1997/1967, p. 21) is not enough, since reports typically employ such sequences as well. More decisive are the expressions of *double temporality*, or the distinction between the time of the telling and the time of the told: narratives are told at a point of time separate from the time the story is set in (Abbott, 2002, pp. 12–19).

The portrayal of emotional experiences has gained significance in the definitions of narrative. Monika Fludernik's (1996, p. 13) theory of narrative maintains that "there can [...] be narratives without plot, but there cannot be any narratives without a human (anthropomorphic) experiencer of some sort on some narrative level". Therefore, for her,

narrativity is constituted by experientiality, that is "the quasi-mimetic evocation of 'real-life experience'" (p. 12). William Labov and Joshua Waletzky (1997/1967, p. 12) also defined narrative as "one method of recapitulating past experience". While Fludernik does not disregard the idea of narrative portraying dynamic movement, she does not understand it as the key element of narrative nor does she want to understand it solely in terms of sequentiality. Still, her model on narrative is based on an episodic structure as an evocation of experience (Fludernik, 1996, pp. 65–66).

David Herman emphasizes that stories constitute a particular loci, which "is an unreplaceable resource for structuring and comprehending experience, a distinctive way of coming to terms with time, process, change." (Herman, 2002, pp. 22–23). Herman's (2009, p. 14) model on narrative takes into account events and experience in the portrayed storyworld as well as the narrating instance. In this model, narrative representations include four basic elements: 1) they occur in a specific occasion of telling and are thus situated in nature; 2) they portray particular events in a structured time-course; 3) the events introduce a disruption in the storyworld; 4) they portray how it feels like for a human or anthropomorphic character to live through the events. Emphasizing the occasion of telling together with the portrayal of events and experience in the storyworld illustrate the importance of double temporality between the time of the telling and the told. Herman's third element, a disruption in the storyworld, associates with the narrative connection, introducing a reason for meaningful resolution.

Based on the theoretical insights above, we determined three necessary conditions for a text segment to be identified as a narrative.

1. The speaker portrays a sequence of events, which are located in another temporal point than the telling (double temporality).
2. These events are situated in a concretely discernible storyworld introducing a tension and a resolution to it (narrative connection and meaningfulness).
3. The events are either a) connected with each other by an experiencing point of view of a character or a group (living through the events) or b) portray a mind other than the speaker's in a manner disclosing the thoughts, feelings and aspirations of another person (evocation of experience). The options for the third criterion allow for several

types of participation and relations the teller might have towards the mediated experience.

Our definition of narrative rests on these three conditions, which are operationalized for the detection algorithm.

The annotation scheme and procedure

The annotation scheme included our conditions for narrative. However, since it is a tool applied interpretatively by the expert annotators, it can also include other categories close to narrative. The task to identify narratives faces the whole fuzziness of the category of "narrative". For example, Fludernik (1996, pp. 57–60) offers a generic overview on oral types of storytelling, mentioning 1) experiential conversational storytelling, 2) the narrative report and 3) the joke and/or the anecdote. In addition to these, there are the cases of "non-spontaneous setting", such as life stories in interview material. Problematically, even reports contain what Labov and Waletzky (1997/1967) call "narrative clauses". J.R. Martin and David Rose (2008, pp. 49–53) similarly distinguish between the (Labovian) narrative, recount, anecdote, exemplum and observation. In order to account for the gradience of narrative, we specified an annotation scheme of four categories: *chronicle*, *narrative*, *almost-narrative* and *narrative-like*. The last two assist with ambiguous yet interesting cases, the former lacking an essential condition of a narrative but otherwise closely resembling one, the latter possessing some properties of a narrative but still not constituting one.

The data was annotated by two annotators, a linguist and a literary narratologist. First, a subset of 6 interviews and 6 plenary sessions was annotated by both of them in order to observe and reach sufficient agreement. In parliamentary records annotator agreement was nearly ideal whereas in interview data slight, but acceptable variation occurred. Next, a random but representative sample of 55 interviews and 60 sessions of the parliament was selected, annotated, and compiled as a sub-corpus for this study.

Documents selected in this article for annotation and computational analysis were chosen across the timespans of both corpora. The selection aimed to preserve a roughly equal

distribution of genders and party allegiances (for interviews) or of different session types (for parliamentary records) but was otherwise random. For parliamentary records, equally many documents were taken from each year. For interviews, two documents were picked from each year: one man and one woman, paying attention to the inclusion of smaller parties in addition to the big ones, and selecting randomly when these factors left several options.

Narrative detection algorithm

Whereas the annotation scheme can build on interpretative conditions of narrative, the algorithm requires specific linguistic features and procedural steps to follow. First, a procedural model was created. To be able to function as a narrative, we postulated that a section of a text must include 1) a nuclear sentence, where the double temporality and narrative meaningfulness are generated and 2) context sentences, where the events and experiences are portrayed in a time-course. Together these types of sentences introduce a speaker portraying a storyworld in another point of time, including event sequences and experiencing subjects. Therefore, the combination of a nuclear sentence followed by at least two context sentences in a sequence is expected to detect narratives. Different context lengths were attempted with expert curation of the results, and two context sentences were found sufficient. The algorithm first searches for nuclear sentences and then attempts to locate context sentences.

To qualify as a nuclear sentence, a sentence must fulfill at least one of the following three criteria designed to capture moments of evoking a telling about past events experienced by the speaker or another person. The alternative linguistic features for a nuclear sentence are 1) a when-clause, 2) a present perfect tense form in the active voice, or 3) a speech act verb in the third person. The when-conjunction was acknowledged as part of our model because it is connected to the expression of episodic or sequential events separate from the time of the telling. The function of when-clauses as recurrent initiators of plotline action in episodic narratives has also been recognized by Fludernik (1996, p. 62).

Since temporal relations between discourse time and denoted past events proved to be contributory factors of narrativity, another accessible frame for our model is the grammatical tense system. In their model on narrative structure, Labov and Waletzky (1997/1967) refer to the tense system in a descriptive manner. They mention that simple past and simple present can function "as grammatical heads" (p. 22) contributing to narrative structure. However, since the simple past abounds in text types such as chronicles, descriptions and arguments, it is not applicable as a nuclear sentence criterion for the algorithm. The Finnish language has a system of four grammatical tenses: present, preterite or simple past, (present) perfect, and past perfect. Preterite is the grammatical device to refer to past events – often described as the narrative tense – and in Finnish is expressed with verb inflection. Perfect and past perfect are periphrastic structures, compound forms of the copula verb "*olla*" and a participle of the main verb. As in many other languages, usage of the perfect tense in Finnish brings denoted past events to the time of the telling with a focus on current relevance (for a more detailed account of the perfect in European languages see Lindstedt, 2000), interrupting the flow of reporting past events with taking a current stance towards the events told (Pallaskallio, 2013), thus creating a narrative connection.

The third feature, speech act verbs, occur in narratively meaningful portrayals of others' aspirations. As will be detailed soon, some of the most common speech act verbs in our data were excluded after testing the algorithm and curating the results. Mental verbs in general are typical of a narrator's portrayal of human experience (see Cohn, 1978), and speech act verbs convey articulated aspirations of self and others. For a context sentence, the sole criterion was a verb in the preterite or the past perfect tense, used to refer to past events.

The algorithm returns sequences of sentences, further referred to as fragments, that are expected to contain narratives. As discussed above, the fragment must necessarily contain a nuclear sentence, followed by at least two context sentences. Context sentences located before the nuclear sentence, or immediately after the two required sentences after it, are added to the fragment. Multiple fragments are joined together if they are consecutive or separated by a single sentence, as most likely they are related to the same candidate narrative. These principles are illustrated in Figure 2.

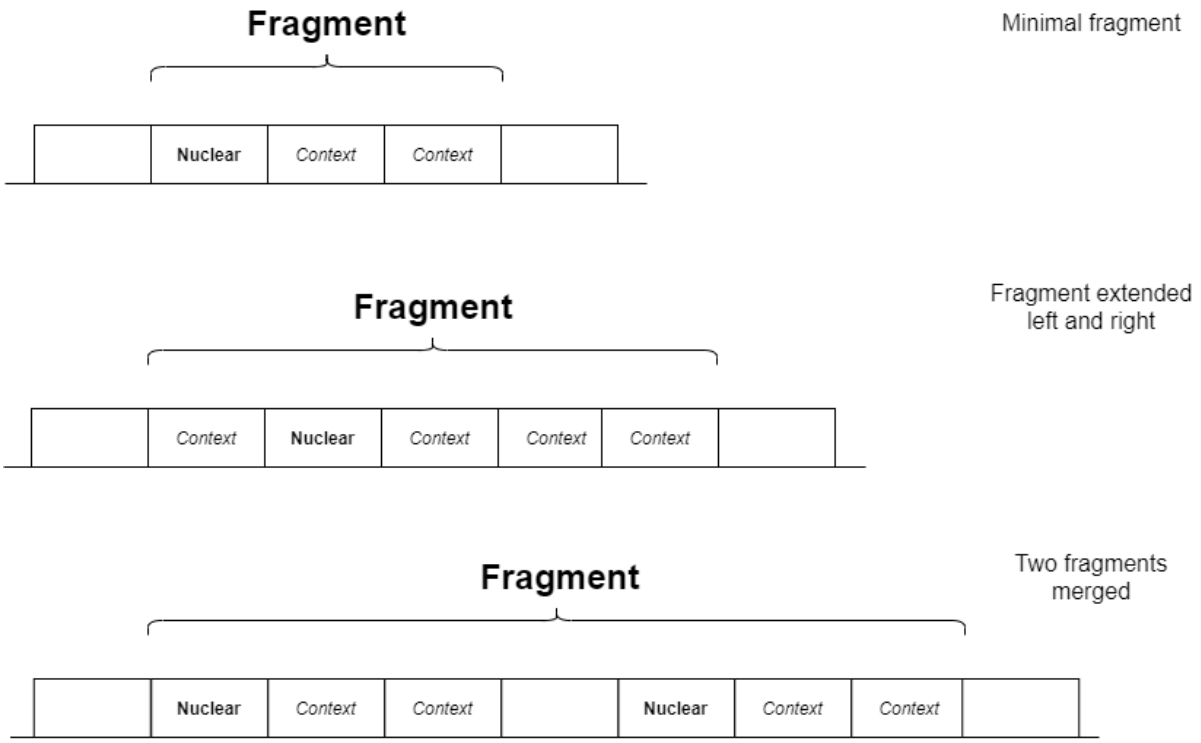

Figure 2. Principles of forming and joining fragments

For the evaluation of the algorithm's performance, both narratives and almost-narratives were treated as correct matches for a narrative. Moreover, any overlap between a fragment and an annotated narrative, even in a single sentence, was counted as a match.

The first versions of the algorithm lacked some of the precision of the current version, indicated in the next paragraph. These versions showed reasonable results when applied to the oral interviews. Within the annotated dataset, they discovered roughly 75 percent of all passages marked as narratives. At the same time, at least half of the algorithm's output consisted of sentences that were not regarded as narrative by the annotators. Performance in the parliamentary records has been substantially worse.

To improve performance, several adjustments have been made to the implementation of the algorithm for the parliamentary records. Firstly, several common speech act verbs, such as "talk" and "speak", were ignored and no longer regarded as nuclear sentence attributes. This is because MPs often use such verbs argumentatively to refer to their colleagues' earlier

statements, without applying them for narrative purposes. Secondly, the perfect tense form identifying nuclear sentences was required to be in the active voice, not the passive. Thirdly, sentences without verbs were marked separately and disregarded when searching for preterite tense forms: the chain of such forms could be interrupted by multiple verbless sentences, as long as at least two preterite sentences were found and no other interruptions occurred. This change accounted for the MPs' formal addressing of the chair, which may occur in the middle of their speech, and other statements that did not conclude the argument.

Finally, sentences that contained at least three preterite verb forms were effectively counted as two context sentences in a row: that is, a single such sentence was sufficient after a nuclear sentence to form a fragment. The reason for this is that sentence boundaries in transcribed speech are somewhat arbitrary, defined by the editing process. Depending on a given MP's speech patterns and pauses made during the talk, a long sentence containing multiple preterite forms could well be broken up into several sentences, which would make them eligible candidates for the narrative detection algorithm.

Quantitative results

The annotators identified a total of 451 narratives in the 55 oral history interviews annotated: the average per document was 8.2 and the median was 7. When applied to the same materials, the detection algorithm yielded 5226 fragments that it classified as narrative. The algorithm correctly recognized 96% of annotated narratives, but it also returned a great amount of fragments that are not narratives. (See also Andrushchenko et al., 2021.) In the 60 plenary sessions, the annotators identified a total of 206 narratives, and the detection algorithm returned 1071 fragments. The algorithm correctly recognized 44% of the narratives, but it also returned an even higher number of other fragments than was the case with the interviews.

The occurrences of narratives and fragments and their overlap are presented in Table 1 along with numerical information on the size of the materials. The first rows describe the outcome

of the annotation, i.e. how many narratives were identified in the documents and how many sentences were marked as belonging to a narrative. The last four rows instead focus on the performance of the algorithm: how many fragments and sentences it yielded, how many fragments did not overlap with annotated narratives and how many narrative sentences appeared in the output.

Table 1. Characteristics of annotated and computationally processed data

Statistic	Interviews	Plenary sessions
Documents	55	60
Sentences	135458	84271
Narratives	451	206
Average narratives per document	8.2	3.4
Median narratives per document	7	2
Narrative sentences	8082	2695
Fragments returned	5226	1071
Empty fragments	4659	974
Narrative sentences returned	6053	442
Total sentences returned	67852	5271

Figures 3 and 4 illustrate the distribution of annotated and computationally retrieved narratives per document, first in interviews and second in plenary sessions. Documents of both datasets are arranged in chronological order. The whole bar represents the total number of annotated narratives, while the black portion corresponds to narratives not picked up by the algorithm.

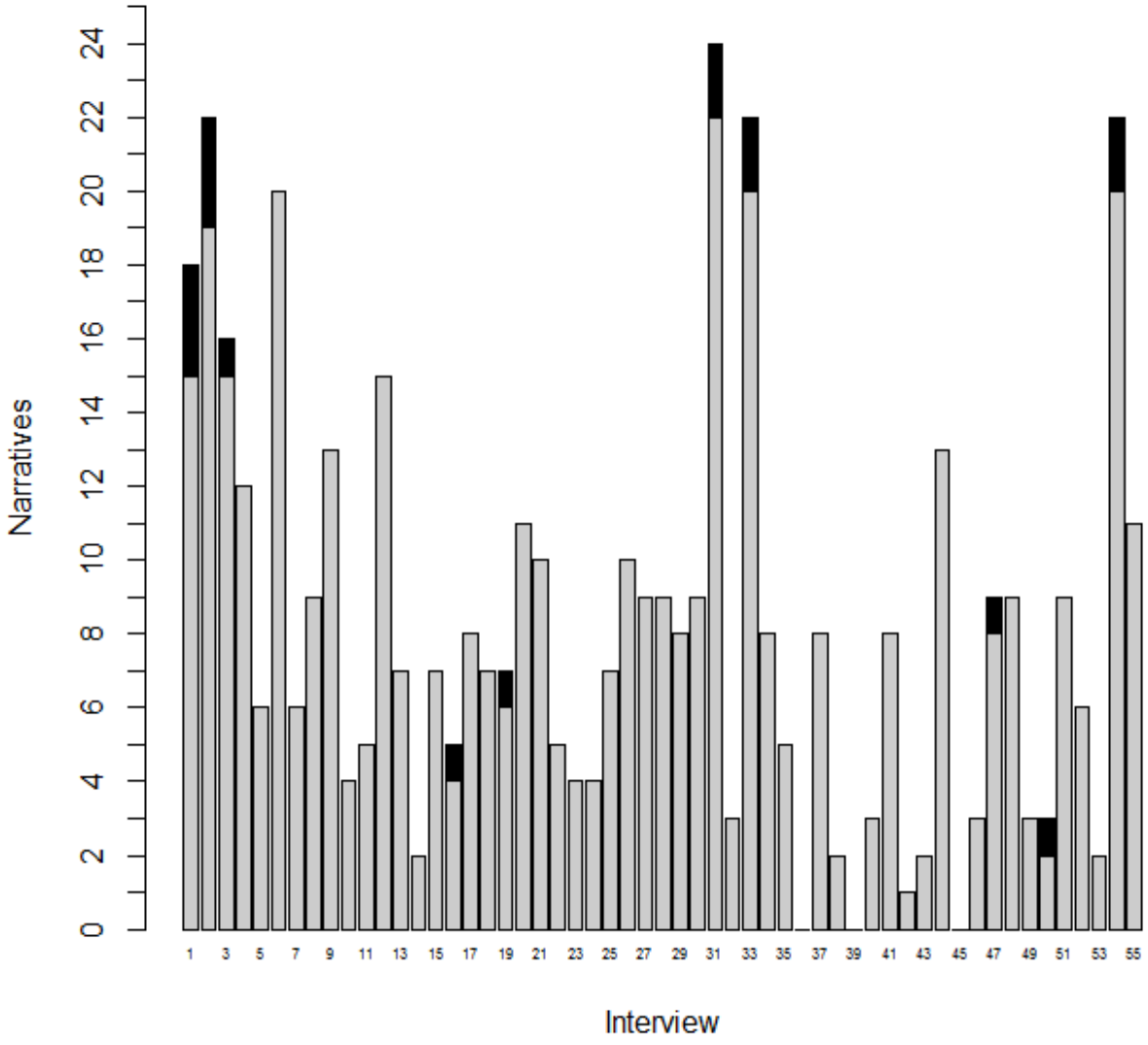

Figure 3. Distribution of annotated and identified narratives in interviews

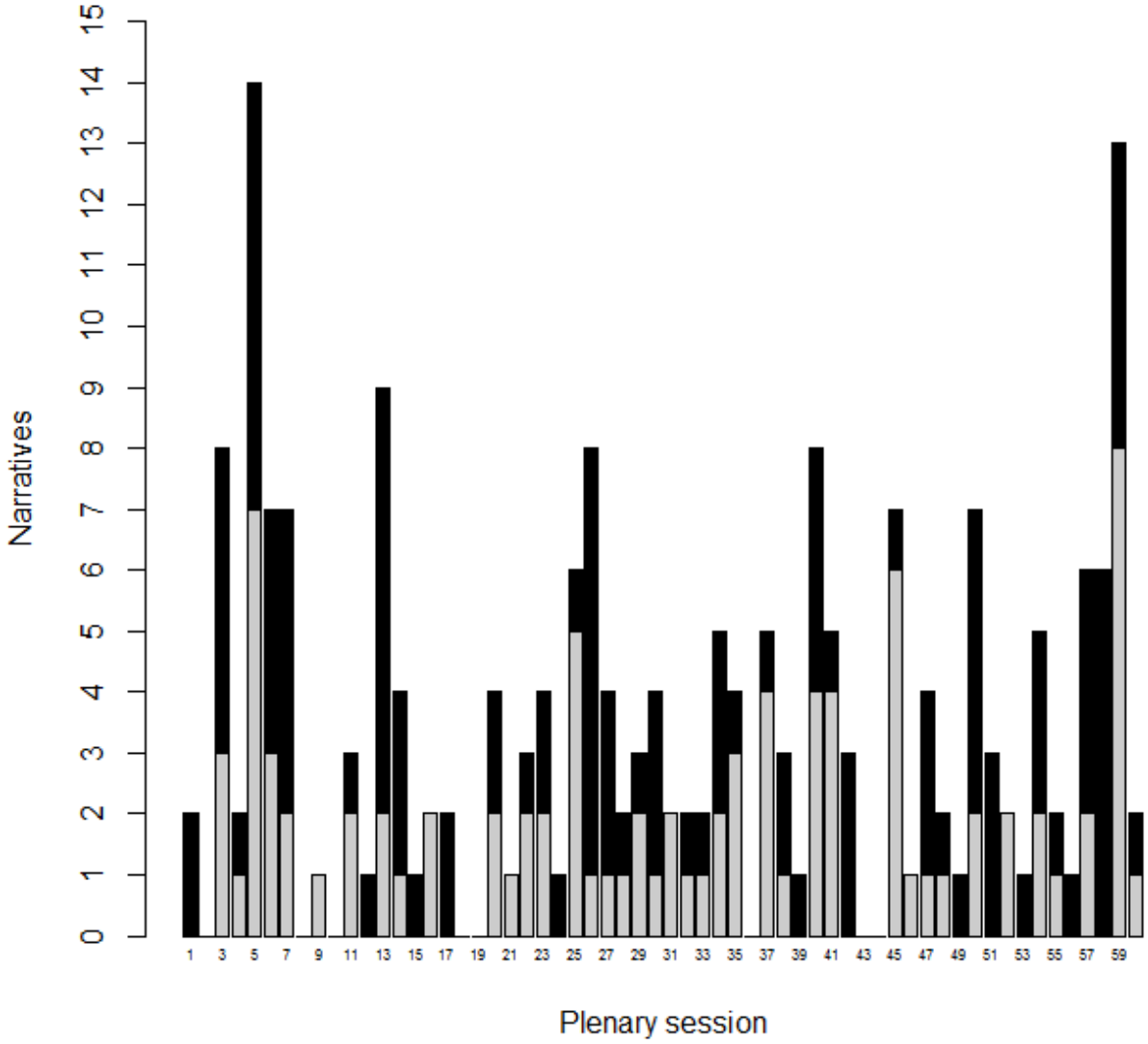

Figure 4. Distribution of annotated and identified narratives in plenary sessions

The figures demonstrate that the algorithm was able to recognize the vast majority of narratives in the interview data, but the success was much poorer in the plenary sessions. What is more, false positives by the algorithm – fragments detected that are not narratives – are also somewhat greater in the plenary sessions than in the interview data: 91% versus 89%.

Since the algorithm's output depends primarily on the nuclear sentences discovered, the rules of their detection have the most impact on the overall performance. These rules, the three criteria of nuclear sentences, do not apply equally often. Table 2 reports how often these criteria contributed to fragments returned by the algorithm. For each fragment, the nuclear sentence that initiated it was counted according to which of the criteria it met. The total sum

is greater than the number of fragments, since some nuclear sentences could satisfy two or three criteria at once.

Table 2. Relative frequencies of nuclear sentence criteria in returned fragments

Nuclear sentence condition	In matches with narratives	Outside of matches	Total
In interviews			
When-clause	462 / 13.66%	2921 / 86.34%	3383
Perfect tense	468 / 12.81%	3185 / 87.19%	3653
Speech act verb	278 / 21.35%	1024 / 78.65%	1302
In plenary sessions			
When-clause	42 / 10.5%	358 / 89.5%	400
Perfect tense	62 / 9.05%	623 / 90.95%	685
Speech act verb	19 / 8.88%	195 / 91.32%	214

The relative frequency of the three elements is somewhat different in the two datasets, which is explainable by their origins and editing conventions. When-clauses, in particular, are somewhat more frequent in interviews than the other elements.

The distribution of the elements between correct and incorrect fragments produced by the algorithm is very close to the ratio of correct and incorrect findings in general. This suggests that none of the nuclear sentence criteria were particularly successful or unsuccessful in identifying actual narratives. An exception is found in the speech act verbs of the interview dataset, which contributed to narrative-matching fragments more often than the other two criteria – and more often than the false positive rate would suggest. More attention to speech act verbs could therefore result in more accurate detection of narratives in oral interviews.

However, the diversity of this component makes the outcome difficult to predict, as it consists of roughly 300 verbs of varying frequency.

Qualitative analysis of the results

Moving to qualitative results, we will first discuss two examples from the interviews, which indicated a high percentage of overlap between the algorithm and the annotators. Then we will move on to discuss several examples of the plenary speeches, where the algorithm both missed a higher percentage of actual narratives and produced a larger number of false positives.

In order to curate the results of the algorithm we needed to apply methods designed to analyze the textual details that evoke the necessary conditions of narrative in each fragment. Earlier we presented our theoretical basis of defining narrative, and in this section we introduce our method for analyzing narratives. This detailed textual analysis also helped to create a feedback loop from the empirical findings to the theoretical definitions in the research process. Fludernik's (1996, pp. 65–70) model, based on story units within a conversational narrative in the work of Labov & Waletzky (1997/1967), enables assessing event formation and the narrative connection. An *abstract* is meant to attract the listener's interest and it introduces the story or invites a storytelling situation. *Orientation* portrays the setting and participants of the storied events. After those introductory parts of a narrative, one or more *episodes* follow, each presenting one or more incidences with a beginning and a result. Episodes are followed by *evaluation* of the point of the story as well as a *coda* about the relevance of the story in a wider context. To analyze the portrayal of experience we utilize Dorrit Cohn's (1978) classical specification on the modes of presenting speech and consciousness in narratives. In Cohn's model, psychonarration denotes instances where the teller portrays a mind of another using their own mental attributions and discursive orientation (Cohn, 1978, pp. 11–12, 31–32). Quoted monologue is a character's speech or thought in their own discourse as if quoted directly (pp. 12, 59–60), whereas a narrated monologue mixes linguistic expressions of the teller and the experiencer (pp. 13, 104–105).

These modes of speech and consciousness representation, originally developed for the analysis of fictional narratives, also help identify and interpret experience portrayed in non-fictional narratives (see Hatavara & Mildorf, 2017).

In the following tables, the second column lists both the grammatical properties picked up by the algorithm (on the first line) and the annotation received by the sentence, if any. The annotation codes are N (narrative), C (chronicle) and AN (almost-narrative). The properties are labelled by i (preterite or imperfect tense), I (multiple preterite forms), p (perfect tense), P (past perfect tense), k (conditional present), K (conditional perfect), s (speech act verb) and W (when-clause); square brackets indicate the boundaries of the fragments returned by the algorithm.

Table 3. Example of agreement between the annotation and the algorithm

Interview 1, 1988

Document line	Props	Original text	Translation
2	[i C	Kysyn ensiksi, miten teistä tuli IKL:n kansanedustaja?	[Interviewer:] I'll first ask: how did you become an MP of the IKL?
3	i C	Se oli oikeastaan varmaan sattuma.	It was actually a coincidence I guess.
4	WPi C	Kekkonenhan oli lakkauttanut IKL:n ja sitten kun tuli vaalit kesällä 39, niin meillä ei ollut järjestökoneistoakaan kunnossa ja kauhea kiire.	Kekkonen had disbanded the IKL and then when the election came in the summer of 39, even our organization didn't click properly and we were in an awful hurry.
5	P	Ja silloin ministeri Eino	And then minister Eino

	C	Tuomivaara, joka oli ollut 14 vuotta eduskunnassa ja kai silloin tällöin ministerikin.	Tuomivaara, who had been in the parliament for 14 years and likely also a minister every now and then.
6	i C	Hän oli IKL:n kansanedustajia.	He was from the IKL's MPs.
7	i C	Hän oli Viipurin läänin itäisessä vaalipiirissä ehdokkaana.	He was a candidate in the eastern constituency of the Viipuri province.
8	Ki C	Tilaston mukaan siitä olisi mennyt läpi vain yksi ainoa edustaja, niin minä lähdin Tuomivaaraa auttamaan.	According to statistics only one single MP would have made it from there, so I went along to help Tuomivaara.
9	i N	Kierreltiin Itä-Suomessa vaalipuheita pitämässä.	We circled across Eastern Finland giving campaign speeches.
10	i N	Ja silloin oli tilanne jo aika uhkaava ja Tuomivaara oli kyllä hyvä puhuja, mutta hän käsitteli maatalouskysymyksiä ja jotakin valtion metsien ja maanviljelijöiden keskinäisien asioiden järjestelyä, että oli sellainen tavallinen enemmän talouspolitiikkaa käsittelevä puhe.	And then the situation was already quite threatening and Tuomivaara sure was a good speaker, but he discussed agricultural questions and some kind of arrangements between the state forests and the farmers, that it was such an ordinary speech that dealt more with financial policy.
11	Pi N	Mutta minä poika, joka olin varamiehenä, niin minä haukuin	But being a hotheaded youngster, who was a deputy, I

		meidän poliitikot, jotka olivat laiminlyöneet armeijan perusteellisesti, että meillä oli vaarauhka siinä.	attacked our politicians who had completely neglected the army, that we faced a threat there.
12	i N	Niin minä pidin aika rajuja puheita ja vapaasti ilman mitään konseptia.	So I gave pretty blunt speeches and freely without any program.
13	Wsi N	Ja minä muistan, että Tuomivaara varoitti minua, että älä pidä niin rajuja puheita, että ukot ei uskalla häntä äänestää, kun apulainen on tuollainen.	And I remember that Tuomivaara warned me: don't make such blunt speeches, that men won't dare to vote for him when his assistant is like that.
14	Wi] N	No sitten kävikin niin, että kun vaalitulokset julkaistiin, niin minä läpäisin ja Tuomivaara tuli varamieheksi.	Well, then it did happen so that, when the election results were published, I came through and Tuomivaara became the deputy.
15	N	Aika nolo juttu.	Quite a silly thing.
16	i N	Minähän olin aivan kokematon varsinaisena virallisena poliitikkona ja niin minä jouduin eduskuntaan.	[Because] I was totally inexperienced as an official proper politician and so I landed in the parliament.

The nuclear sentence here is number 4 thanks to the when-clause. The next sentence contains a past perfect form instead of the preterite, allowing the algorithm to collect the remainder of the fragment up to sentence 14. Sentences 2–8 were annotated as a chronicle and the rest as a narrative.

This narrative begins with an abstract on the question and continues on line 3, which both intrigues the listener and portrays the teller as a humble person (containing thus an element of evaluation). Lines 4 to 9 provide orientation about the national political situation and at the

end the events of the story to come. Lines 10 to 12 provide an iterative narrative of what happened many times – the two men giving different types of speeches – whereas line 13 builds up towards the result through a prediction then proved utterly wrong in line 14, which gives the surprising resolution of the narrative. Line 15 provides the evaluation both within the storyworld and from the time of the telling, pinpointing the embarrassment for the main candidate, and line 16 offers a coda, introducing the interviewee starting as a rookie in the post he is being interviewed about. This narrative neatly follows the model for episodic narratives.

The narrative also provides access to the minds of the two persons in the storyworld. The teller describes himself and Tuomivaara as different types of speakers. Line 13 provides a mixture of quoted speech and narrated speech from Tuomivaara, and an ironic stance towards him is created by the narrator mockingly quoting this schematized warning, which in line 14 turns out to be based on a complete misunderstanding of the electorate's sentiments. The partial appropriation of the character's past speech serves to evoke past experience.

Table 4. A chronicle that the algorithm identified as a narrative

Interview 1, 1988

Document line	Props	Original text	Translation
34	[ki	Jos vielä palattaisiin siihen itse tavallaan vaalikiertueeseen, niin miten se tapahtui noin aivan konkreettisesti, liikuitteko autolla?	[Interviewer:] If we still returned to the, kind of, electoral tour itself, then how did it happen quite concretely, did you travel by car?
35	i C	Me liikuimme autolla ja meidän piirisihteeri, joka oli Sulo Laine.	We travelled by car and our district secretary, who was Sulo Laine.
36	Pi	Hän oli entinen punakaartilainen ja	He was a former red guard and a

	C	sosialisti ja oikein reipas poika, niin oli tuota -18 ollut Viipurissa panssarijunan konduktöörinä.	socialist and a really lively lad, had driven an armored train in Viipuri in -18.
37	i C	Hän järjesti nämä kiertueet ja väkeä oli tuvan täydeltä.	He arranged these tours and the place was packed to the rafters.
38	i C	Mutta meidän organisaatio ja raha-asiat ja muut oli hyvin heikot, mutta viipurilaiset sentään järjestivät rahaa.	But our organization and financial matters and others were really weak, but still the people of Viipuri came up with money.
39	i C	Minä muistaakseni sain tohtori Toivo Sepältä, joka oli varakas lääkäri ja oli niin sanottu piirinjohtaja, tosin Viipurissa, niin saatiin 5000 markkaa, millä se kierros käytiin.	As I remember, I received from doctor Toivo Seppä, who was a wealthy doctor and was a so-called head of district, though in Viipuri, so got 5000 marks that were spent on the tour.
40	i C	Siihen aikaan rahan arvo oli vähän toinen.	The value of money was a bit different at that time.
41	i C	Ja Laineella oli jonkinlainen vanha autonrämä, millä sitten kierrettiin.	And Laine had some ramshackle old car that we drove around.
42	i C	Oli kuljettaja mukana.	[Interviewer:] Had a driver along.
43	i C	Laine itse ajoi.	Laine drove himself.
44	p C	Hän on kai kuollut.	He must have died.
45	pi AN	Olen nähnyt myös tällaisia julisteita, joita tehtiin tai oliko se	[Interviewer:] I have also seen such posters that were made, or

		lehti-ilmoitus, missä vetoatte aika voimakkaasti kansaan.	was it a newspaper advert, where you appeal to the people quite forcefully.
46	i AN	Ne oli julisteita.	Those were posters.
47	i] AN	Se minun vaalijuliste, se minun lankoni Imatralla, tuomari Zitting järjesti sen ja maksoi sen painatuksen ja se poikkesi normaalista vaalikehumisesta julisteissa täydellisesti ja sitä pantiin joka ikiseen telefonitolppaan ja muuhun ja se taisi olla aika tehokasta.	That election poster of mine, that my brother-in-law in Imatra, judge Zitting arranged it and paid for its printing and it differed completely from the usual election praise in posters and it was slapped onto every single telephone pole and such and it was possibly quite effective.
48	k AN	Suosittelisin vastaisuudessakin politiikoille vähän terhakkaampia vaalijuttuja, niin se voisi auttaa.	In the future I'd also recommend a bit more lively election talk to politicians, that could help.

The detection algorithm has arrived at this sequence from sentence 44, containing a perfect tense form, which is both followed and preceded by sentences with preterite verbs. However, the answer to the interviewer's first question was annotated as a chronicle and the last three sentences as the start of an almost-narrative. On lines 35–37, one could even locate a glimpse of politically intensive narrativity. It introduces the district secretary, who had been on the left side in the Finnish Civil War of 1918 but later a very active district secretary and driver of the far right party IKL (Patriotic People's Movement). Here the evaluation, the point of the story, is historically so obvious that the narrator does not feel the need to mention it. In the interview dataset, the continuum from "chronicles" when the speaker recounted a sequence of events, without attempting to tell a complete narrative, to proper narratives posed a challenge. Fludernik (1996, p. 62) already notices that oral history interviews tend to invite such long

sequences of events since the speaker does not have the conversational pressure to be concise and relevant.

We will next move to examples from the parliamentary sessions.

Table 5. Agreement between the annotation and the algorithm

Session 17, 2005

Document line	Props	Original text	Translation
1052	p N	Perintö- ja lahjavero on yksi sellainen vero, joka on tullut tiensä päähän suomalaisen yhteiskunnan kannalta nykyisessä globaalissa maailmassa.	The tax on inheritance and gifts is one such tax that has run its course with respect to the Finnish society in today's global world.
1053	p N	Täällä salissa on tullut ilmi monia hyviä perusteita perintö- ja lahjaveron poistamiseksi.	Many good grounds for the abolishment of inheritance and gift tax have emerged in this chamber.
1054	N	Kysymys on siitä, että verottaja käy uudelleen kimppuun jo kertaalleen verotettuun omaisuuteen.	The problem is that the taxman strikes again at property already taxed once.
1055	N	Tässä on samanlaisia piirteitä kuin luonnossa tapahtuu, että karhu, nyt verokarhu, käy haaskalla.	This has similarities to what happens in nature, when the bear, now the taxman bear, comes at bait.
1056	[i N	Yksi esimerkki verottajan kovasta kohtelusta kävi ilmi tänään.	One example of the taxman's rough treatment emerged today.

1057	PI N	Sain sähköpostiviestin, jossa palaute oli, että kuolinpesän pieni metsäomaisuus oli jouduttu myymään, jotta veroista voitiin selvitä.	I received an email where the message was that the small forest share of the estate had had to be sold in order to cope with the taxes.
1058	WI N	Myyntitulot olivat kuitenkin aikataululla, johon verottaja ei voinut suostua, ja tältä osin ainut mahdollisuus oli vekselin teko, kun ei ollut käteistä rahaa.	However, the sale profits followed a schedule that the taxman could not agree to, and in this regard the only option was to make a loan contract, since there was no cash available.
1059	i N	Lisäksi perinnönsaaja taisteli jo ennestään asuntolainan kanssa, ja tilanne tämän surun keskellä ei kyllä varmasti mieltä ylentänyt.	Moreover, the inheritor was already struggling with a mortgage from before, and the situation, in the midst of this grief, certainly didn’t raise spirits.
1060	I] N	Totean vielä, että perikunta haki verottajalta lykkäystä verojen maksuun, mutta se aikataulu ei tyydyttänyt verottajaa, ja perikunta joutui ottamaan velkaa veron maksamista varten.	I’ll also mention that the inheritors applied for an extension of the payment deadline from the tax service, yet the schedule did not satisfy the taxman, and the inheritors were forced to take loans in order to pay the tax.

The last five sentences of the excerpt were picked up by the algorithm: the nuclear sentence was 1058, featuring a when-clause, and preterite verbs were found in the following two sentences and also in the preceding two. However, the narrative sequence begins from line

1056, which functions as an abstract for the story explicitly stating that an example of the topic is to follow. The three preceding lines provide necessary context information to establish narrative connection. Line 1057 provides an orientation, giving the basic situation and further summarizing the story to come. Lines 1058 to 1060, also picked up by the algorithm, provide episodes on what has happened and thus portray the story events. Line 1059 includes an evaluation, as the sad effects of the accumulating hardships are given. This evaluation is already foreshadowed in lines 1054 to 1055, and the coda, the relevance of the story, is alluded already in the first sentence of the narrative on line 1052.

This narrative contains all the elements of an episodic narrative. Moreover, line 1059 provides an indirect and supposed mental state of the protagonist regarding their spirits and grief. These are portrayals of their emotional experience in the form of psychonarration. The metaphor of the taxman as a bear and the reference to the taxman's rough treatment further emphasize this experiential weight of the narrative, exemplifying the unfairness of the inheritance tax.

Table 6. Annotated narrative missed by the algorithm

Session 85, 2014

Document line	Props	Original text	Translation
1187	N	Asia, johon haluan puuttua, on väliinputoajien valitettava joukko.	The matter I want to focus on is the unfortunate group of those who drop through the net.
1188	N	Etenen tässä yhden yksittäisen tapauksen kautta, joka valottaa tätä ongelmaa.	I'll proceed here with one particular case that illustrates this problem.
1189	i N	Nuori nainen, 24-vuotias, menetti työpaikkansa tuotantotaloudellisista syistä.	A young woman, 24 years old, lost her job for production and economy-related reasons.

1190	Pi N	Oli ollut ahkera työntekijä, mutta menetti työpaikkansa.	Had been a hard-working employee, but lost her job.
1191	I N	Te-toimiston kannustamana hän haki ja pääsi avoimeen korkeakouluun opiskelemaan, ja tavoitteena oli tietysti parantaa työllistymismahdollisuuksia.	Encouraged by the unemployment office, she applied for and was admitted to open university studies, and the goal was of course to improve her employment opportunities.
1192	I N	Opiskelupaikan saaminen ja vastaanottaminen kuitenkin johtivat siihen, että te-toimisto eväsi työttömyyskorvauksen, koska opiskelu katsottiin päätoimiseksi.	However, receiving and accepting the study right caused the unemployment office to deny the unemployment benefits, since studies were seen as the primary occupation.
1193	ik N	No, ei mitään hätää, nuori nainen päätti ottaa yhteyttä Kelaan, jotta saisi opintotukea.	Well, no worries at all, the young woman decided to contact Kela [the social security service] to receive study benefits.
1194	i N	Kelan mielestä hänen opiskelunsa oli sivutoimista, ja sen takia opintotukea ei voitu myöntää.	Kela had the opinion that her studies were a part-time occupation, and therefore the study aid could not be granted.
1195	 N	(Puhemies koputtaa) Onko tämä ongelma tiedossa, ja mitä suosittelette tässä tilanteessa olevalle nuorelle naiselle?	(Chair taps gavel) Are you aware of this problem, and what do you recommend to the young woman who is in this situation?

Here the chain of preterite verbs is not sufficient for the algorithm to identify the passage as a narrative. Line 1190 exhibits the past perfect tense, not the perfect, and is treated as the preterite, and there are no speech act verbs or when-clauses either.

Lines 1187 to 1188 provide the abstract, and line 1189 the orientation, the former inviting the storytelling situation and the latter giving the setting for the story to start. Lines 1190 to 1194 provide several episodes that build up to and illustrate the events with a clear narrative connection between them. The evaluation is not articulated directly, but the coda on line 1195 implies it: the story is an example of the ill-treatment of those the system does not recognize, thus the protagonist is ill-treated.

This narrative presents the mind of the character, the anonymous young woman, in many ways. Particularly notable are lines 1193 and 1194. In the former, the character's thoughts are portrayed via quoted interior monologue ("no worries at all"), which gives the impression of a cheerful, persistent person with faith in things getting better despite some obstacles. Her determination to improve her life is portrayed using a mental verb typical of psychonarration ("decided"). The cheery expectation and decisive action is in the following line 1194 juxtaposed with what the officials thought, building a stark contrast between the young woman, who had been characterized as "hard-working" (line 1190) and manifested grit, and the anonymous collective of the agency denying support to her. This agency is portrayed as a subject with a collective mind hosting an opinion. Therefore, the vicarious minds are depicted in this narrative with opposite interests and views, which emphasizes the experiential relevance of the narrative and the sequence of complicating actions. Our detection algorithm, however, is not sophisticated enough to recognize shifts from one mind to another without precise linguistic features that could be included in it.

Table 7. A fragment mistakenly identified as narrative by the algorithm

Session 117, 2005

Document line	Props	Original text	Translation

529	[Wp	Kun sovittelun vahvistavaa lakia on vaatimalla vaadittu tässä eduskuntasalissa, se on pysähtynyt nimenomaan hallintokuntien, oikeusministeriön ja sosiaali- ja terveysministeriön, näkymättömään lasiseinään, jonka… (Välihuuto) — No, valtiovarainministeriökin on voinut olla siinä, mutta kyllä myös tämä yhteistyö on ollut auttamattoman heikkoa.	When the law reinforcing the arbitration was demanded here in the parliamentary chamber, it has stopped precisely against the government committees', the ministry of justice's and the ministry of social affairs and health's invisible glass wall, which... (Interjection) — Well, the ministry of finance, too, may have been there, but indeed also this cooperation has been hopelessly weak.
530	i	— Kaksi vuotta sitten jouluna saimme valtiovarainministeriön kanssa yhteistyössä budjetin käsittelyssä hyväksytyksi varat sovitteluun.	— Two years ago on Christmas we reached an agreement with the ministry of finance on the funds in the budget proposal.
531	WI]	Niitä varoja ei voitu lisätä budjettiin, kun ei ollut lakia, ja ministeri ei suostunut antamaan lakia, kun ei ollut vielä budjetissa rahoja.	Those funds could not be added to the budget, since there was no law, and the minister did not agree to issue the law, since there were no funds in the budget yet.

The first sentence of the passage contains a when-clause and a perfect tense form, and preterite verbs appear in the following sentences. The content of this fragment does include episodes in several points of time, for example two years ago. The perfect tense on line 529 suggests that the situation described is still ongoing, which creates a link from the time of the telling to the time of the told. Still, the fragment does not portray an experiencing subject or events with narrative connection contributing to each other or a resolution.

Table 8. Another mistakenly identified fragment

Session 129, 1999

Document line	Props	Original text	Translation
4132	[i	Vuonna 1991 kuntien palveluksessa oli 432 000 työntekijää.	In the year 1991 municipalities employed 432 000 workers.
4133	i	Vuonna 1998 kuntien palveluksessa oli 416 000 työntekijää eli noin 15 000 työntekijää vähemmän kuin vuonna 91.	In the year 1998 municipalities employed 416 000 workers, that is, about 15 000 fewer employees than in the year 91.
4134	ps	Tänä aikana on toteutettu muun muassa subjektiivinen päivähoito alle kouluikäisille, joka on vaatinut melkoisia työntekijälisäyksiä, ja ne ovat noissa luvuissa siis mukana.	During this time, in particular, subjective daycare has been implemented for preschoolers, which has demanded substantial employee increases, and they are accordingly included in those figures.
4135	i	Viime talvena ilmestyi Oecd-maista tehty työn kiireen tutkimus, jota on tehty muistaakseni 70-luvun loppupuolelta saakka.	Last winter a study of feeling time pressure in OECD countries was published, which has been conducted, as I recall, since the second half of the 70s.
4136	i]	70—80-luvun alussa kaikkein kiireellisimmän työn kuvaukset tulivat naisvaltaisista tehdasympäristöistä vaihetyöntekijöiltä.	In the 70s and the early 80s, descriptions of the most pressured work came from production line workers in female-dominated factory

			environments.

The algorithm arrived at this fragment starting from sentence 4134, finding two context sentences ahead of it and then extending to the two behind it. Notably, the verb “demand” is used here in a different connotation, without referring to an actual speech act. Still, its tense ensures that the sentence qualifies as a nuclear sentence. The perfect here does not include a stance towards the past events, nor do the events portrayed in the context sentences suggest episodic structure with narrative connection between them.

Conclusion

In this study, we succeeded in developing a relatively reliable computational method for recognizing narratives in two globally non-narrative large datasets. What is more, our approach enables distinguishing narratives from chronicle-type portrayal of past events. The algorithm introduced in the study allows for a more systematic investigation into the use of narratives in the Finnish parliament, as well as in other properly digitized Finnish-language datasets. This enables multiple applications in all disciplines working with large textual data. Since the algorithm correctly identified most of the narratives in interviews, and a large portion of those in parliamentary records, it can be used as a greatly time-saving initial selection, from which any false positives can be removed as postprocessing by experts. Their interpretative effort can then be quickly directed to those automatically identified cases of interest.

The results are definition-bound in the sense that, for example, the study of habitual or hypothetical narratives would require adjustments of the algorithm. Still, this approach, where a rule-based algorithm was tested and curated by expert annotation, demonstrates that the study of such highly interpretative and contested humanities and social sciences phenomena as narratives can be aided by automated analysis of big data. Quantitatively, the accuracy is good for the interview materials and satisfactory for the talk in the parliament, particularly taking into consideration the complex nature of the phenomenon detected.

Qualitatively, some challenges can be identified in spite of the overall performance. First, it seems challenging to automatically detect narratives where changes between presented minds and voices are not marked; the listener understands these shifts, but the algorithm cannot operate on world knowledge or contextual information. Second, from the talk in the parliament the algorithm sometimes falsely returns fractions of event portrayal that lack sufficient narrative connection but report several procedural happenings at different points of time.

Since narrative is a key resource for mediating experience and making sense of time and change, the study of narratives across time and narrative environments is crucial for any discipline working with human action. Future work includes applying and modifying our model to analyze other large data sets both within and beyond political talk. The results of our research project indicate a possibility to profoundly change both the scope of historical analysis and our understanding of the occurrences and role of narratives in different societal spheres. Computational recognition of the key narrative passages enables targeting the interpretative effort of humanities and social sciences experts without having to do all labor-intensive reading manually. It also reveals the situations and contexts where narratives occur thus enabling the study of circumstances that more often than others call for narratives to be employed. What is more, the iterative cycles between theoretical definitions, annotation and reading the results of the algorithm may help to locate unexpected, emergent phenomena within narratives and to estimate relations between different features of narrative.

The findings obtained in this work are partially applicable to materials in other languages. They confirm the assumption that the complex system of grammatical tenses functions together with semantic elements of, for example, word choices in relating past events. The components of the narrative detection algorithm – the preterite and the perfect tense, the when-clause and the category of speech act verbs – are expected to perform roughly the same high-level functions across languages. Our model, acting as a tentative mapping of these properties to narrative features, is therefore extendable rather than restricted only to Finnish. Moreover, the semantic classification of verbs could be extended outside of speech act verbs to study the affordances of using, for example, mental verbs or emotive verbs as part of the model. However, the resulting performance relies on whether operationalized lexico-

grammatical features are indeed related to narrative phenomena, as the results of this study suggest.

References

Abbott, H. P. (2002). *The Cambridge introduction to narrative*. Cambridge University Press.

Andrade, S. B., & Andersen, D. (2020). Digital story grammar: A quantitative methodology for narrative analysis. *International Journal of Social Research Methodology,* 23, 405-421. https://doi.org/10.1080/13645579.2020.1723205

Andrushchenko, M., Sandberg, K., Turunen, R., Marjanen, J., Hatavara, M., Kurunmäki, J., Nummenmaa, T., Hyvärinen, M., Teräs, K., Peltonen, J., & Nummenmaa, J. (2021). Using parsed and annotated corpora to analyze parliamentarians' talk in Finland. Journal of the Association for Information Science and Technology, 73(2), 288–302. https://doi.org/10.1002/asi.24500

Bögel, T., Strötgen, J., & Gertz, M. (2015). A hybrid approach to extract temporal signals from narratives. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology* (pp. 106-107). https://konvens.org/proceedings/2015/GSCL-201500.pdf

Carroll, N. (2001). On the narrative connection. In W. van Peer & S. Chatman (Eds.), *New perspectives on narrative perspective* (pp. 21-41). State University of New York Press.

Cohn, D. (1978). *Transparent minds: Narrative modes for presenting consciousness in fiction*. Princeton University Press.

Eisenberg, J. D., & Finlayson, M. A. (2016). Automatic identification of narrative diegesis and point of view. In *Proceedings of 2nd Workshop on Computing News Storylines* (pp. 36-46). Association for Computational Linguistics. http://doi.org/10.18653/v1/W16-5705

Ek, A., & Wiren, M. (2019). Distinguishing narration and speech in prose fiction dialogues. In C. Navaretta, M. Agirrezabal, & B. Maegaard (Eds.), *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference* (pp. 124-132). CEUR-WS.org. CEUR-WS.org/Vol-2364/

Fludernik, M. (1996). *Towards a 'natural' narratology*. Routledge.

Genette, G. (1980). *Narrative discourse revisited*. Cornell University Press.

Hatavara, M., & Mildorf, J. (2017). Fictionality, narrative models, and vicarious storytelling. *Style*, 51, 391-408.

Haverinen, K., Nyblom, J., Viljanen, T., Laippala, V., Kohonen, S., Missilä, A., Ojala, S., Salakoski, T., & Ginter, F. (2014). Building the essential resources for Finnish: The Turku dependency treebank. *Language Resources and Evaluation*, *48*(3), 493-531. https://doi.org/10.1007/s10579-013-9244-1

Herman, D. (2002). *Story logic: Problems and possibilities of narrative.* Nebraska University Press.

Herman, D. (2009). *Basic elements of narrative*. Wiley-Blackwell.

Labov, W., & Waletzky, J. (1997/1967). Narrative analysis: Oral versions of personal experience. *Journal of Narrative and Life History,* 7(1-4), 3-38. (Reprinted from *Essays on the verbal and visual arts: proceedings of the 1966 annual spring meeting of the American Ethnological Society* ed. by June Helm, Seattle: University of Washington Press, 1967).

Lin, C., Wright-Bettner, K., Miller, T., Bethard, S., Dligach, D., Palmer, M., Martin, J. H., & Savova, G. (2020). Defining and learning refined temporal relations in the clinical narrative. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis* (pp. 104-114). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.louhi-1.12

Lindstedt, J. (2000). The perfect – aspectual, temporal and evidential. In Ö. Dahl (Ed.), *Tense and aspect in the languages of Europe* (pp. 259-277). Mouton de Gruyter.

Mani, I. (2014). Computational narratology. In P. Hühn, J. C. Meister, J. Pier, & W. Schmid (Eds.), *Handbook of narratology* (pp. 84-92). De Gruyter, Inc.

Martin, J. R., & Rose, D. (2008). *Genre relations. Mapping culture*. Equinox.

Miller, B., & Park, J. S. (2020). Computing narrative. In F. Karsdorp, B. McGillivray, A. Nerghes, & M. Wevers (Eds.), *Proceedings of the Workshop on Computational Humanities Research* (pp. 182-190). CEUR-WS.org. http://ceur-ws.org/Vol-2723/

Pallaskallio, R. (2013). *Kertova tempus: finiittiverbin aikamuodon valinta suomenkielisissä katastrofiuutisteksteissä 1860–2004* [Narrative usage of tense in Finnish disaster news texts 1860–2004]. University of Helsinki, Faculty of Arts, Department of Finnish, Finno-Ugrian and Scandinavian Studies. http://hdl.handle.net/10138/40335

Piper, A., So, R. J., & Bamman, D. (2021). Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on EMNLP* (pp. 298-311). Association for Computational Linguistics. https//doi.org/10.18653/v1/2021.emnlp-main.26

Prince, G. (1982). *Narratology. The form and functioning of narrative*. Mouton.

Prince, G. (1987). *A dictionary of narratology*. University of Nebraska Press.

Rappaport Hovav, M., & Levin, B. (2015). The syntax-semantics interface: Semantic roles and syntactic arguments. In S. Lappin & C. Fox (Eds.), *The handbook of contemporary semantic theory* (pp. 593-624). 2nd ed. Wiley Blackwell. https://doi.org/10.1002/9781118882139.ch19

Reiter, N., Willand, M., & Gius, E. (2019). A shared task for the digital humanities chapter 1: Introduction to annotation, narrative levels and shared tasks. *Journal of Cultural Analytics* 4(3). https://doi.org/10.22148/16.048

Sagae, K., Gordon, A. S., Dehghani, M., Metke, M., Kim, J. S., Gimbel, S. I., Tipper, C., Kaplan, J., & Immordino-Yang, M. H. (2013). A Data-Driven Approach for Classification of Subjectivity in Personal Narratives. In M. A. Finlayson, B. Fisseni, B. Löwe, & J. C. Meister (Eds.), *Workshop on Computational Models of Narrative 2013* (pp. 198–213). Schloss Dagstuhl – Leibniz-Zentrum für Informatik. http://doi.org/10.4230/OASIcs.CMN.2013.198

Sudhahar, S., Franzosi, R., & Cristianini, N. (2011). Automating quantitative narrative analysis of news data. In T. Diethe, J. Balcázar, J. Shawe-Taylor, & C. Tîrnăucă (Eds.), *JMLR: Workshop and Conference Proceedings* 17, 63-71. http://proceedings.mlr.press/v17/sudhahar11a/sudhahar11a.pdf

Tammi, P. (2006). Against narrative. A boring story. *Partial Answers*, 4(2), 19-40.

Virtanen, T. (1992). Issues of text typology: Narrative – a 'basic' type of text? *Text*, 12, 293-310.

VoDe Corpora. (2021). Parliamentary text corpus. Voices of Democracy Project, Plenary sessions of the parliament of Finland from 1980 to 2021. Interview corpus. Voices of Democracy Project, 404 veteran parliamentarians' interviews. Both corpora grammatically parsed. Requires separate access rights.

Voutilainen, E. R. J. (2017). The regulation of linguistic quality in the official speech-to-text reports of the Finnish parliament. *CoMe: Studies on Communication and Linguistic and Cultural Mediation*, 61-73. http://comejournal.com/wp-content/uploads/2019/06/5.-CoMe-II-1-2017.-Voutilainen.pdf

White, H. (1987). *The content of the form. Narrative discourse and historical representation*. The Johns Hopkins University Press.

Yang, H., & Garibaldi, J. M. (2015). Automatic detection of protected health information from clinic narratives. *Journal of Biomedical Informatics*, 58, S30-S38. https://doi.org/10.1016/j.jbi.2015.06.015

Zhang, H., Boons, F., & Batista-Navarro, R. (2019). Whose story is it anyway? Automatic extraction of accounts from news articles. *Information Processing and Management*, 56(5), 1837-1848. https://doi.org/10.1016/j.ipm.2019.02.012

Addresses for correspondence:

Mari Hatavara
address: Tampere University, 33014 Tampere University, Finland
e-mail: mari.hatavara@tuni.fi
Orcid: 0000-0003-2970-2770

Kirsi Sandberg
address: Tampere University, 33014 Tampere University, Finland
kirsi.sandberg@tuni.fi
Orcid: 0000-0002-3682-1375

Mykola Andrushchenko
address: Tampere University, Kalevantie 4, 33014 Tampere University, Finland
e-mail: mykola.andrushchenko@tuni.fi

Sari Hälikkö

address: Tampere University, Kalevantie 4, 33014 Tampere University, Finland
e-mail: sari.halikko@tuni.fi
Orcid: 0009-0004-1606-4209

Jyrki Nummenmaa
address: Tampere University, 30014 Tampere University, Finland
e-mail: jyrki.nummenmaa@tuni.fi
Orcid: 0000-0002-7476-7840

Timo Nummenmaa
address: Tampere University, 33014 Tampere University, Finland
e-mail: timo.nummenmaa@tuni.fi
Orcid: 0000-0002-9896-0338

Jaakko Peltonen
address: Tampere University, 33014 Tampere University, Finland
email: jaakko.peltonen@tuni.fi
Orcid: 0000-0003-3485-8585

Matti Hyvärinen
address: Tampere University, 33014 Tampere University, Finland
e-mail: matti.hyvarinen@tuni.fi
Orcid: 0000-0002-1145-9656