Chapter 2

Formality in mediated and non-mediated discourse: Bringing together human judgements and corpus-driven detection

Ilmari Ivaska[a], Adriano Ferraresi[b] & Marta Kajzer-Wietrzny[c]
[a]University of Turku [b]University of Bologna [c]Adam Mickiewicz University

Several works have suggested that both interpreting and translation tend to favour linguistic choices typically considered more formal. Observations concerning the degree of formality of these forms of language mediation, however, have been made within studies mostly focusing on different phenomena, such as standardization or conventionalization. In this paper, we report on a quantitative and qualitative analysis focusing specifically on formality in interpreted language. We take into account native and interpreted speeches delivered at the European Parliament (EP) and collected in the English subcorpora of the EPIC and EPTIC corpora. We compare the examined English varieties to one another, taking into account the mode of delivery of the speeches, i.e. whether they were read out from a written text or delivered impromptu. Our hypothesis is that interpretations of speeches read at the EP are located at the far end of the formality spectrum compared to speeches delivered impromptu by native English MEPs. Unlike previous work, we rely on formality indicators identified by triangulating human judgements and specific linguistic features derived bottom-up from a corpus, based on the MuPDAR[F] approach. The analysis provides partial support for the hypothesis, showing that interpreted texts are generally predicted as being characterised by a high level of formality, irrespective of the actual mode of delivery of their source text. We conclude by commenting on the linguistic features that were found to contribute the most to making interpreted speeches diverge from native ones.

Ilmari Ivaska, Adriano Ferraresi & Marta Kajzer-Wietrzny

1 Introduction

As pointed out by Heylighen & Dewaele (1999), all speakers are likely to intuitively distinguish between formal and informal registers, whereby the prototypical formal end of the spectrum resembles the language used by a judge during a trial, and the prototypical informal end is marked by a relaxed conversation among friends. Graesser et al. (2014: 218) associate formality with the need to be "precise, coherent, articulate, and convincing to an educated audience", as opposed to informal settings such as oral conversation, which is "narrative, replete with embodiment words" and reliant on common background knowledge. Andrén et al. (2010: 224) link formality with the application of "officially standardized and recognized institutional conventions or prescriptions".

Although research on formality per se in Interpreting and Translation Studies is still scarce, a lot of attention has been directed towards the related notions of standardization (Toury 1995) and conventionalization (Baker 1993). Thus, it has been suggested that translations tend to shun informal language use, e.g. by relying on "generally unmarked grammar clichés, and typical, common lexis instead of the unusual or the unique" (Mauranen 2008: 41). In turn, this tendency may be related to risk-aversion (Pym 2005), whereby, to transfer the source meaning, translators, and by extension interpreters, opt for conventional linguistic forms.

In this paper we report on a quantitative and qualitative analysis of formality of native and interpreted speeches delivered at the European Parliament and collected in the English subcorpora of the European Parliament Interpreting Corpus, or EPIC (Sandrelli et al. 2010), and the European Parliament Translation and Interpreting Corpus, or EPTIC (Ferraresi & Bernardini 2019). In the reported study we compare the examined English varieties to one another, further taking into account the mode of delivery of the speeches, i.e. whether they were read out from a written text or delivered impromptu. Our working hypothesis is that interpretations of speeches read at the European Parliament are located at the far end of the formality spectrum compared to speeches delivered impromptu by native English MEPs.

Our approach to analyzing formality has been inspired by Jarvis' work on lexical diversity, particularly on the observation that while lexical diversity is an emic construct, in that "it relies crucially on the human interpretation of both form and meaning" (Jarvis 2017: 540), it has been traditionally studied following an etic approach, "concentrat[ing] on the identification, measurement, and description of forms and their distribution in a way that does not rely on meaning" (ibid.). In line with Jarvis' thinking, we find that formality, too, is a profoundly emic construct, and so any potential automated measurement should be based

on a holistic, human-informed definition of the construct (Jarvis 2013a; Jarvis 2013b; Jarvis 2017). Hence, in the reported study we rely on formality indicators identified through a data-driven, human-informed operationalization of linguistic formality.

We start with an overview of studies touching on issues of formality in interpreting and translation (§2), and then move on to a detailed description of the method that allowed us first, to identify the linguistic features characteristic of texts that have been classified as formal by humans (§3), and then to discuss our results (§4). In §5, we conclude by summarizing our main results and suggesting ways in which the results of the present study could be validated and applied.

2 Formality in Interpreting and Translation Studies

To the best of our knowledge, formality has so far not been the focal point of any analysis in Interpreting and Translation Studies, but rather a feature commented on in the context of studies on interpreting and translation investigating other linguistic phenomena. Since formality has so far merited only a mention in these papers, we decided to include both interpreting and translation in our literature overview to gain a better perspective on formality as a phenomenon in mediated discourse in general. This will be used as a backdrop for our findings regarding formality specifically in interpreting.

While no studies focus explicitly on formality in simultaneous interpreting, several reports mention features pointing to greater formality in this type of communication. Kajzer-Wietrzny (2012) compares the frequency of the optional connective *that* in translations and interpretations into English from Romance and Germanic languages, as well as spoken and written native English speeches. Findings point to the connective being more frequent in both spoken and written mediated texts. As *that*-omission might be related to informality (Biber 1995: 145), higher frequencies of the optional connective can be interpreted as a manifestation of formality. Moreover, in an interview preceding a case study of interpreting style, one of the interpreters reported using more formal language and higher register in his interpretations than in his non-interpreted communication (Kajzer-Wietrzny 2012).

From a completely different perspective, in an analysis of face-threatening acts during plenaries at the European Parliament, Bartłomiejczyk (2016) points to cases where interpreters made the target rendition less face-threatening and "considerably more formal by consistently avoiding any colloquial vocabulary" (Bartłomiejczyk 2016: 203) and by "using more formal and euphemistic language"

(Bartłomiejczyk 2016: 211). It follows that formality in interpreting might also be a by-product of strategies mitigating impoliteness.

Hale (1997: 57) found that interpreters working at court hearings, instead of approximating the register of the speaker, "by and large adapt their communicative style to what they perceive to be the expectations and/or limitations of the listener", raising or lowering the level of formality depending on the interpreting direction. She reports that the level of formality is raised by strategies such as the condensation of the source text and the omission of typically conversational features such as fillers, hesitations, repetitions and backtracking. On the other hand, the level of formality is lowered especially through lexical choices, e.g. translating a formal word with a colloquial item, or adding semantically empty words or phrases of pragmatic importance signalling "a certain level of familiarity" (Hale 1997: 47–50). Such changes of communicative style are key "in forming impressions", which in the context of court interpreting has an impact on the perception of the witness in court (Hale 1997: 53).

Another observation, potentially pertinent to the issue of formality in interpreting, comes from a study carried out by Shlesinger (1989). She observed an equalizing effect in interpreting, whereby "the interpretation of texts which exhibited typically literate features" tended to be "shifted towards the oral end of the continuum, whereas the interpretation of texts that exhibited typically oral features shifted towards the literate end" (Shlesinger 1989 in Shlesinger & Ordan 2012: 54). Assuming that oral vs. literate features correspond to various degrees of formality, such findings stress the need to further investigate formality in interpreting.

Preference for formality also emerges from studies on written translation. In her study on the occurrence of contractions in literary translation and contemporary literary English writing, Olohan (2003) found differences "in terms of both variety of contracted forms encountered and frequency of occurrence of contractions" (Olohan 2003: 59); together with *that*-omissions, these can be seen as a "crude measure of informality" (Olohan 2004: 101). She observes variation between translators, which might depend on the source language, level of formality of the source text, translator's style, and/or genre and narrative structure of the text. It also follows from the discussion of results that texts in the analysed translation corpus display more features typical of Biber's (1988) informational production, i.e. they are less involved, less generalised, more explicit and more edited.

Translations have also been found to display a more "unmarked formal register and a neutral standard variety of the language" (Zlatnar Moe n.d.: 125). Looking at various types of register shifts, Zlatnar Moe (n.d.: 136) concludes that translators

"shift the style from the extremes towards neutrality", and that shifts towards increased formality constitute the largest number of shifts in all analysed texts. Such shifts of register, she argues, may make the text less appealing for the reader.

One of the few works that have looked closely at formality in translation is the study by De Sutter et al. (2012: 343), who use profile-based correspondence analysis and logistic regression to assess whether translations use more formal language than other texts. The first method allows the authors to measure and visualize the linguistic distances between the language varieties. Logistic regression makes it then possible to evaluate "the exact impact of the lects on the lexical choices" (De Sutter et al. 2012: 325). The authors select 10 lexical variables, consisting of a neutral and a formal variant of semantically equivalent expressions, and use them to explore formality distances between texts of different genres originally written in Dutch, and Dutch translations from either French or English. They conclude that translated texts and non-translated texts differ with respect to formality, but also that translations are not uniform: translations from French are more formal than those from English. Furthermore, text type turns out to be the most important variable. Interestingly, text types differ with respect to formality, but "[t]here appears to be no formality differences between translated and non-translated" texts within specific text types, such as journalistic texts, non-fiction and instructions (De Sutter et al. 2012: 340). It must be noted though, that according to the authors "formality variation cannot be predicted successfully" on the basis of text types and source languages only, since the reported regression model explains 18% of the variation.

Comments on formality also emerge from studies on constrained language, where translation is frequently set against native and non-native language varieties. Kruger & van Rooy (2018: 237) observe that "non-native varieties and translated texts avoid informality features in written registers" and relate it to the authors' level of language proficiency. Less proficient users are more prone to risk avoidance. On the other hand, the tendency to increased formality is one of the shared features of two constrained varieties, i.e. translated and non-native indigenised varieties of English that distinguish them from the native variety (Kruger & van Rooy 2016: 26). This transpires from several shared tendencies, such as a greater mean word length than in the native varieties, lower frequency of the pronoun *it*, less frequent use of emphatics (e.g. *really*, *for sure*, *a lot*), avoidance of possibility modals (e.g. *can*, *could*, *may*, *might*), avoidance of stranded prepositions, higher frequency of nominalizations and increased frequency of *wh*-questions (Kruger & van Rooy 2016: 37–41). Other features pointing to increased formality are observed in translation only (not in non-native varieties), such as the use of optional connective *that* or higher frequency of nouns.

As this overview should have shown, many side-remarks are found in various studies which suggest that both interpreting and translation seem to favour linguistic choices typically considered more formal. With a few exceptions (e.g. De Sutter et al. 2012), these indications still need to be tested in a study devoted specifically to formality.

3 Data and method

To test whether interpreting is more formal than non-mediated native spoken discourse, a general operationalization of formality is needed. Following in part the methodological architecture of Jarvis (Jarvis 2017: 548–549), we adopt a study design combining corpus-derived and human-informed data. Specifically, focusing on speeches delivered at the European Parliament as a case in point, we: 1) ask human judges to evaluate a set of non-interpreted, native data in terms of their perceived overall formality; 2) train statistical classifiers on a different set of comparable non-interpreted native data to evaluate their formality, using a range of potentially relevant, linguistically defined featuresets; 3) use the trained classifiers to predict the formality of the human-evaluated data, and zoom in on the linguistic features that contribute most to the successful classification; 4) use this final model to analyse a set of comparable interpreted texts, so as to assess how they are positioned with respect to the non-interpreted texts in terms of formality features. Each of these steps and the associated datasets are described in the following sections.

3.1 Data

3.1.1 The European Parliament (Translation and) Interpreting Corpora

The corpus data analyzed in this study come from two related corpora, i.e. the European Parliament Interpreting Corpus (EPIC, Sandrelli et al. 2010) and the European Parliament Translation and Interpreting Corpus (Ferraresi & Bernardini 2019). Both of them are multilingual corpora of speeches given in the plenary sessions of the European Parliament, but while EPIC only includes transcriptions of original speeches and their interpretations, EPTIC also features the corresponding written-up versions of the original speeches (so called "verbatim reports") and their written translations. The languages currently represented in the corpora, as sources, targets or both of interpreted and translated texts, are English, Italian and Spanish (EPIC), and English, French, Italian, Polish and Slovene (EPTIC), for a total of around 580,000 tokens.

All the data used in this study are in English, and are extracted from the spoken component of the corpora, comprised of non-interpreted, original speeches (henceforth called "original"), and interpreted ones. The dataset can be divided into three subsets: 1) native original train data, 2) native original test data, 3) interpreted data with French, Italian, and Polish as source languages. Speeches were selected so as to provide a balanced dataset with respect to their mode of delivery, i.e. whether they were originally delivered impromptu or read out, a distinction which we hypothesize to be associated with formality differences (see §3.2.1). It should be noticed that in the case of interpreted texts, the mode of delivery refers to the original speech.

The transcripts follow the audio recordings, but given that the purpose of the present paper is to detect and analyse characteristic features of formality and not general differences between spoken and written language, we have excluded from the transcripts elements that are exclusive to spoken language, including hesitations, false starts, as well as empty and filled pauses (editing guidelines can be found in Appendix A). As suggested by one reviewer, such orality features might contribute to the perception of formality, yet their elimination was necessary so as to avoid excessive weight being assigned to them both in the human and the automatic evaluation task.

The texts were then parsed according to the Universal Dependencies (UD) annotation scheme using the Turku neural parser (Kanerva et al. 2018), and the subsequent analyses were conducted using the parsed data. Table 1 gives an overview of the corpus data used.

Table 1: Analysed dataset

	Native train	Native test	fr>en	Interpreted it>en	pl>en
Impromptu	30 texts	10 texts	10 texts	5 texts	10 texts
	6,573 w	1,797 w	1,940 w	821 w	1,607 w
Read out	30 texts	10 texts	10 texts	5 texts	10 texts
	7,566 w	2,045 w	1,825 w	912 w	1,506 w
All	60 texts	20 texts		100 texts	
	14,139 w	3,842 w		8,611 w	

3.2 Method

3.2.1 Rationale and research questions: grounding formality in human perception

The definitions of formality have often been functional or contextual. By way of example, the notion of "formal" is defined by Atkinson (1982) as the opposite of "conversational" and by Andrén et al. (2010) as adherence to institutional conventions, while Heylighen & Dewaele (2002) refer to high and low contexts (for a detailed discussion, see Li et al. 2016). Methodologically, it is noteworthy that definitions of the theoretical construct of formality are often emic in nature, in that they rely on "human interpretation of both form and meaning" (Jarvis 2017: 540), holistically taking into account both function and form.

The operationalizations of formality, however, have traditionally been linked exclusively to form. Several scores based on linguistic indicators have been proposed to measure formality in texts, including the formality score, which takes into account the frequencies of various word classes (Heylighen & Dewaele 2002), the adjective density score (Fang & Cao 2009), or lists of formal and informal words created ad hoc (Abu Sheikha & Inkpen 2010). Within corpus-driven approaches, formality has also been used as an explanatory tool when interpreting differences across texts from different categories (e.g. registers, genres or varieties). For instance, in the wealth of studies making use of multidimensional analysis, formality is often used to explain the nature of linguistic differences observed between registers (e.g. Biber 1988; Conrad & Biber 2001; Biber 2012). In other words, it is interpreted as the reason why certain categories diverge from one another, as attested by differences in frequencies of linguistic elements. In both scenarios, the link between formality and the linguistic features that attest to it is an indirect one, inasmuch as assessments of (the degree of) formality have typically not been rooted in human perception: in Jarvis' terms, they are etic operationalizations.

In this paper, we explore a data-driven and human-informed operationalization of linguistic formality for European Parliament data. In the first part of the analysis, we take as our point of departure two text classes that, we hypothesize, are characterized by different levels of formality, i.e. speeches delivered impromptu (less formal) vs. read out (more formal). We then establish their order of perceived formality by means of human judgements, and use these judgements as a gold standard in a data-driven analysis of linguistic features distinguishing the more formal vs. less formal text class. As we are interested in potential differences between interpreted and non-interpreted communication, we explore a

range of linguistic features that have been used earlier to distinguish mediated from non-mediated language use (Volansky et al. 2015).

Our specific research questions are: 1) Are read out speeches perceived as more formal than impromptu speeches? 2) Which linguistic features contribute to distinguishing these text classes? 3) Do interpretations differ from spoken native non-interpreted texts in terms of formality, as assessed by the linguistic features thus identified?

3.2.2 Human evaluations

Our aim was to follow the emic approach and the example set by Li et al. (2016), and so we wanted to root our text-based operationalization of formality – the distinction between speeches delivered impromptu and those read out loud – on human judgements. More specifically, we wanted to assess whether, and to what degree, human evaluators agree on formality differences between texts when they are not given any linguistic definition of formality itself.

In total, 55 individuals participated in an online survey where they were shown ten pairs of texts and were asked to choose for each pair which text was more formal in their opinion. There was one impromptu text and one read out text in each pair, and to minimize any test effect, the pairs were assigned randomly, so that each impromptu text was paired at least once with every read out text, and vice versa. The order in which the texts were introduced, as well as their order in the display, changed randomly for each participant. Participants were not given any definition of formality in the beginning, but they were told where the texts stem from. At the end of the survey, they were also asked to elaborate on their notion of formality (the questionnaire form can be found in Appendix B). The survey was circulated among the authors' colleagues and students, and not advertised in any other way. It was conducted anonymously on the Webropol online survey platform (version 3.0[1]) in 2019. The texts evaluated in this way constitute the test set of our corpus data.

According to the background information provided, the first language of 53 participants out of 55 is other than English. Altogether, 48 participants reported having a higher education background in languages, linguistics, translation and/ or interpretating, and all but two agreed that they feel comfortable studying or working in English. The age range of informants were: 18–25 years (27), 26–35 years (14), 36–45 years (10) and more than 45 years (4). One participant did not complete the survey, and their answers were left out of the results.

---

[1]https://webropol.com

3.2.3 Featuresets considered

Our ultimate research question was to see whether interpreted language diverges from non-interpreted language in terms of formality, and to explore the linguistic features that contribute to this potential difference. Hence, we decided to compare a range of different featuresets which have been shown to consistently distinguish mediated from non-mediated language in general, and to assess which ones could also be related to formality differences.[2] As all the included featuresets were implemented using parsed and CONLL-U formatted data,[3] they are easily transferrable to data on any language where sufficient UD resources are available.

Various kinds of sequential n-grams are probably the most widely used featuresets (e.g. Baroni & Bernardini 2006; Koppel & Ordan 2011; Volansky et al. 2015). We, too, used normalized frequencies of unigrams, bigrams and trigrams of words, lemmas, parts-of-speech (POS), as well as syntactic functions, for a total of 12 featuresets. For instance, example (1) consists of six unigrams (*I*, *want*, *to*, *ask*, *some*, *questions*), five bigrams (*I want*, *want to*, *to ask*, *ask some*, *some question*) and four trigrams (*I want to*, *want to ask*, *to ask some*, *ask some questions*), each of which is represented in the four different levels of annotation (word, lemma, POS, syntax).

(1)

I	want	to	ask	some	questions	(word)
I	WANT	TO	ASK	SOME	QUESTION	(lemma)
PRON	VERB	PART	VERB	DET	NOUN	(POS)
nsubj	root	mark	xcomp	det	obj	(syntax)

Positional tokens, i.e. starts and ends of sentences, are other features that have been used to distinguish mediated from non-mediated texts (e.g. Volansky et al. 2015; Rabinovich et al. 2016). To that end, we considered first, second, penultimate and ultimate positions in sentences to see how often different items occur in these positions. Here, too, we used four parallel featuresets, one for each level of annotation. In the case of example (1), we looked at how often *I/I/PRON/nsubj* occurred in the first position of the sentence, *want/WANT/VERB/root* in the second position, and so on. The sentence boundaries stem from the original data structure of EPTIC.

---

[2] All the frequency data as well as the R scripts of the statistical analyses can be found here: https://osf.io/q75jw/

[3] https://universaldependencies.org/format.html

Character trigrams and other character n-grams have been shown to work well when distinguishing mediated from non-mediated texts (e.g. Popescu 2011; Volansky et al. 2015). However, as pointed out by Volansky et al. (2015: 113), character-based features are difficult to interpret in a linguistically meaningful manner, especially when shorter n-grams are considered. Due to this difficulty, we opted for a compromise solution which allowed us to include this feature without totally sacrificing interpretability: we decided to focus exclusively on the trigram level and to limit the focus to individual words and their boundaries. For instance, the sequence *I want* consists of the character trigrams *_I_*, *_wa*, *wan*, *ant* and *nt_*.

More recently, dependency bigrams have been introduced as reliable, scalable and yet linguistically interpretable features when distinguishing mediated from non-mediated texts (Ivaska & Bernardini 2020; Ivaska et al. 2022). Unlike typical POS bigrams, dependency bigrams are not necessarily sequential and provide information on the constituent words/lemmas/POS, their order in text, as well the nature of the syntactic relation linking them.
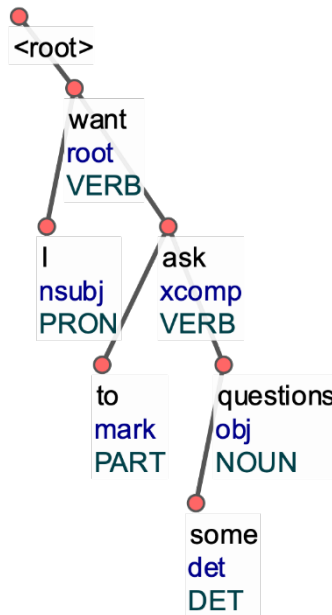

Figure 1: Tree representation of example 1

For instance, example (1) (visualized in Figure 1) consists of the sentence *I want to ask some questions*, split into the following dependency word bigrams: *I*NODE_nsubj_*want*HEAD, *want*HEAD_xcomp_*ask*NODE, *to*NODE_mark_*ask*HEAD,

*ask*HEAD_obj_*questions*NODE, *some*NODE_det_*questions*HEAD. We used three dependency bigram featuresets, defined on the word level, the lemma level, and the POS level.

3.2.4 Feature selection and model training

For each of the 20 featuresets considered in the train dataset, we conducted a feature selection procedure. First, following the example of Volansky et al. (2015), we limited each featureset to the 300 most common features. Then, to tease apart those features that actually contribute to the classification task on text formality, we conducted a Boruta feature selection (Kursa & Rudnicki 2010) for each featureset. According to Kursa & Rudnicki (2010), feature selection is helpful in predictive model building, as modern datasets are frequently rich in irrelevant variables that may decrease models' accuracy. Hence selecting a "small (possibly minimal) feature set giving best possible classification results is desirable for practical reasons" (ibid). Boruta helps to limit the dataset to only the most relevant variables.

Boruta introduces randomness to the data by duplicating all variables and randomly permuting the duplicates' values (here, feature frequencies). It then makes use of the random forest algorithm (Breiman 2001) and builds a classification model for the task at hand (here, the identification of texts presumably characterized by different formality), compares the actual features' performance to the randomized features, and suggests as important only those features that consistently outperform the randomized duplicates. Random forest was chosen as the statistical method, as it was originally created to solve issues related to data including few observations with many predictors (ibid.), much like ours. The sizes of our final featuresets are summarized in Table 2.[4]

We then trained separate forest-based classifiers for each featureset using the same train dataset consisting of impromptu and read out files (labelled *read_-001*, *impromptu_001* etc.) but only the selected features as predictors. We used the *ranger* implementation of random forests throughout the analyses (Wright & Ziegler 2017) and trained a prediction model for each featureset. We trained the models with ranger's probability function to obtain the likelihood of each prediction instead of just the most likely label. Whenever the classification is discussed in terms of the predicted labels, we have used 0.5 as the cut-off point.

---

[4]Due to the relatively small train dataset, the word trigram featureset ended up being too sparse for reliable feature selection, and it has thus been left out of all the subsequent analyses.

Table 2: The featuresets used in the formality classification task

Category	Featureset	Final number of features (out of considered)
Sequential n-gram	word 1-gram	12 (of 300)
	lemma 1-gram	14 (of 300)
	POS 1-gram	3 (of 15)
	syntax 1-gram	7 (of 40)
	sequential word 2-gram	11 (of 300)
	sequential lemma 2-gram	6 (of 300)
	sequential POS 2-gram	11 (of 187)
	sequential syntax 2-gram	13 (of 300)
	sequential word 3-gram	NA
	sequential lemma 3-gram	6 (of 300)
	sequential POS 3-gram	12 (of 300)
	sequential syntax 3-gram	14 (of 300)
Positional frequencies	positional word	7 (of 300)
	positional lemma	7 (of 300)
	positional POS	10 (of 55)
	positional syntax	11 (of 97)
Character n-gram	character 3-gram	18 (of 300)
Dependency n-gram	dependency word 2-gram	7 (of 300)
	dependency lemma 2-gram	9 (of 300)
	dependency POS 2-gram	9 (of 300)

3.2.5 Model validation and MuPDAR[F] analysis

The rest of the analysis follows the logic of the Multifactorial Prediction and Deviation Analysis using Regression / Random Forests (MuPDAR[F], e.g. Gries & Deshors 2014; Gries & Adelman 2014 using regression; e.g. Deshors & Gries 2016; Gries & Deshors 2020 using random forests), a two-phase analysis where a certain phenomenon (in our case: original mode of delivery) is modelled in train data and, provided that the model predicts the phenomenon well, the same model is used to predict the phenomenon in different data that diverge from the train data in some respect. In our case, original models are trained on non-interpreted data and they are used to predict interpreted data. Analysing the deviations oc-

curring in the predictions on the second dataset, based on the model trained with the first dataset, gives a detailed insight on the ways in which the two datasets diverge from one another. The method has been used successfully when explaining how L1 and L2 users of English or L2 users with different L1 backgrounds diverge from each other (e.g. Gries & Deshors 2015 on dative alternation differences between EFL and ESL learners; Wulff & Gries 2019 on L1-related variation in verb–particle constructions in L2 English), but also to contrast translated with non-translated English (Kruger & De Sutter 2018 on *that*-omission).

In the present study, we used the obtained forest models of each featureset to predict the original mode of delivery in the test data. We then selected the model of the best-performing featureset and used that to predict the mediated data. Looking at the direction of the deviation provides an overall view on the role of formality in mediated language use: if texts delivered originally impromptu are predicted more often falsely as having been read out loud than the other way around, the results can be seen to indicate that the mediated texts are indeed relatively more formal than non-mediated ones. Comparing the results of the machine prediction with the human judgements (cf. §3.2.2) makes it possible to validate (or reject) the applicability of the different featuresets as indicators of a formality difference.

As a final step, MuPDARF logic allows for further analysis of the observed deviations. To this end, we built a final forest model on the erroneously predicted mediated data. Here, we followed the logic of Deshors & Gries (2016): we had as the response a numeric variable that indicated how far off the prediction was from being correct. On the other hand, the predictors were the features included in the final model, as well as the constraining language. The values of the variable range from -0.5 to 0.5, where negative values represent cases where a read out text was predicted as impromptu, and positive values cases impromptu text was predicted as read out. The further away from zero the value, the more erroneous the prediction.

In short, we first trained a range of forest-based classifiers to distinguish impromptu speeches from those read out in the non-mediated native English variety; these models involved a categorical response variable (impromptu vs. read out). We then evaluated which of the featuresets make it possible to fit the most accurate model. Having selected the most viable model for the native English non-mediated speeches, in line with the MuPDARF approach, we trained another model to predict to what extent mediated/interpreted speeches deviate from the outcomes of the native English non-mediated variety. In the latter case the response variable was a numeric one, i.e. the observed deviation.

4 Results

4.1 Human evaluations of formality differences

Table 3 shows for each of the 20 texts how many times (out of 54) each read out text was labelled as more formal, and each impromptu text as less formal than the other text in the same (shuffled) pair.

Table 3: Perceived formality of read out and impromptu texts

Text ID	Subjects labelling text as more formal	Text ID	Subjects labelling text as less formal
read_05	53 (98.1%)	impromptu_07	51 (94.4%)
read_04	49 (90.7%)	impromptu_05	50 (92.6%)
read_09	49 (90.7%	impromptu_01	49 (90.7%)
read_01	47 (87.0%)	impromptu_08	49 (90.7%)
read_10	47 (87.0%)	impromptu_06	46 (85.2%)
read_08	45 (83.3%)	impromptu_02	45 (83.3%)
read_02	44 (81.5%)	impromptu_03	44 (81.5%)
read_03	44 (81.5%)	impromptu_09	44 (81.5%)
read_07	41 (75.9%)	impromptu_10	41 (75.9%)
read_06	38 (70.4%)	impromptu_04	35 (64.8%)

In all cases, more than half of the respondents perceived read out texts as more formal and impromptu texts as less formal, with percentages of agreement among raters above 80% for 16 texts out of 20, equally split across the two categories.

53 out of 55 respondents further provided answers to the final question of the survey asking them to illustrate the reasons for their choices regarding the more formal text. These answers were categorized bottom-up to gain an understanding of the linguistic or stylistic features that respondents associated with formality or informality. Table 4 reports on these categories, with percentages indicating the proportion of answers mentioning them, as well as examples of the specific features mentioned by respondents.

In §4.2.1 we compare and discuss some of these features, especially those concerning lexis and syntax, with respect to those emerging from the text-based analysis of formality differences.

Table 4: Features associated with text formality

Category	Features (e.g.)
Discourse (35.7%)	Coherence/cohesion
	Impersonal style
	Lack of repetitions
Lexis (32.2%)	Rare vocabulary
	Terminology
	Formulae (e.g. greetings)
Syntax (28.0%)	Complex sentence structure
	Sentence length
	Passive forms
Content (4.1%)	Hard facts (vs. opinions)

4.2 Corpus-based identification of formality differences

4.2.1 Best-predicting features of formality differences

The 20 trained forest-based classifiers were used to predict formality differences within the test data, corresponding to the 20 texts which were also used in the human evaluation experiment. The performance of each classifier was evaluated in terms of precision, i.e. the probability that the model classifies texts correctly as read out or impromptu, and recall, i.e. the ability of the model to find all instances of read out and impromptu texts. Table 5 reports precision and recall values alongside the resulting F-measure (the harmonic mean of the other two values), which is conventionally used to assess the accuracy of classification models. For space reasons, results are only reported for the 10 best-scoring models.

With 6 featuresets among the overall top 10, the category of sequential n-grams seems to perform better than other categories, and especially n-grams of length 1, i.e. unigrams (of syntactic functions, POS and words). Syntax-based features also perform well, both when used as part of sequential n-grams and within positional-based featuresets. Among the top 3 models, dependency POS bigrams get the highest F-measure together with syntax 1-grams, while at the same time packing more linguistic information than the other two best-scoring featuresets (i.e. unigrams of syntactic functions and POS respectively), since they a) take into account syntactic functions in the form of dependencies, and b) are based on POS.

Table 5: Classification accuracy of the 10 best-scoring models

Featureset	Precision	Recall	F-measure
syntax 1-gram	0.8	0.8	0.8
dependency POS 2-gram	0.8	0.8	0.8
POS 1-gram	0.7	0.875	0.778
positional POS 1-gram	0.76	0.716	0.737
word 1-gram	0.8	0.667	0.727
sequential syntax 3-gram	0.8	0.571	0.667
sequential syntax 2-gram	0.7	0.636	0.667
sequential POS 2-gram	0.6	0.667	0.632
positional syntax	0.58	0.658	0.616
positional lemma	0.6	0.545	0.571

In view of the small differences in terms of F-measure as well as the higher linguistic and functional interpretability of the featureset, which was also observed in previous work (cf. §3.2.4), we selected the model based on dependency-defined POS bigrams as the most suitable classifier for subsequent analyses.

Nine features (Table 6) contribute to the distinction between read out and impromptu texts in our analysis. The five features more frequently observed in texts originally read out, which we hypothesized to be characterized by greater formality, all involve nouns. In particular, three of these involve a determiner used in conjunction with a noun, where the noun is typically related to the topic of the discussion, and the determiner is used to increase precision (e.g. *all exports*, *the threat*, *some enlargements*). Proper nouns serving as a direct object of a verb are often used in the formula of expressing thanks (*thank Iñigo*), and nouns coordinated by a conjunction are frequently observed in the formula opening a speech (*ladies and gentlemen*). The latter feature is also observed in excerpts studded with terminology (e.g. *scrutiny and control* or *deficit and debt*). Altogether, these observations tie in both with previous literature and with comments made by respondents in our survey. Heylighen & Dewaele (1999) noted that nouns and noun phrases are typically more frequent in formal texts, and both formulaic expressions and use of specialized terminology were noted by respondents as also associated with increased formality.

The remaining features are more frequent in impromptu texts, which we relate to informality. Three of them involve verbs, and specifically verbs having pronouns as subjects (*we took*) and verbs associated to adjectives or nouns (*sus-*

Table 6: Features distinguishing impromptu and read out texts illustrated by examples

POS dependency bigram	Example	Register where more frequent
DETNODE_det.predet_-NOUNHEAD (determiner predetermining noun)	I believe Thailand should be completely delisted from all poultry meat exports until they can prove they have the infrastructure	Read out
DETNODE_det_NOUNHEAD (determiner determining noun)	The threat to Europe's health from the rapid spread of disease is real and present.	Read out
VERBHEAD_obj_PROPNNODE (verb having a proper noun as a direct object)	Let me thank Íñigo and the coordinators for taking the decision	Read out
NOUNHEAD_conj_NOUNNODE (nouns connected by coordinating conjunction)	I genuinely wonder, ladies and gentlemen, where the human rights regaining of confidence of the markets is the basis for a stable, sustainable growth and jobs	Read out
DETHEAD_nmod_NOUNNODE (determiner modifying noun)	many of them raise concerns about some of the previous enlargements	Read out
NOUNHEAD_advmod_ADVNODE (adverb acting as modifier of noun)	that's the agricultural sector here in Europe that room over there	Impromptu
PRONNODE_nsubj_VERBHEAD (pronoun acting as subject to verb)	It is time we took action and showed that we support the Iranian opposition	Impromptu
ADJHEAD_conj_VERBNODE (adjective coordinated with verb)	Now this is not sustainable, and it's not fair	Impromptu
NOUNHEAD_advcl_VERBNODE (noun acting as adverbial clause modifier of verb)	It's a very good initiative but the Parliament, as Mr de Jong says, has some concerns.	Impromptu

tainable [...] it's; *initiative [...] says*); the fourth one involves adverbs modifying nouns (*sector here*). These features are consistent with Heylighen & Dewaele's observation that verbal structures tend to be preferred in informal contexts to nominal structures (1999), and that informal style is often associated with deictic expressions, which are exemplified here by pronouns (*we took*) and adverbs such as *there* or *here*. Inflected verbs, too, are suggested by Heylighen & Dewaele (1999) to be "intrinsically deictic because they refer implicitly to a particular time through their tense (...), and to a particular subject through their inflection". Highly context-dependent, deictic expressions are bound to decrease precision and increase involvement, which renders the text more informal.

The next section reports on the results of the MuPDAR[F] analysis based on the model featuring the dependency bigrams discussed here.

4.2.2 MuPDAR[F] analysis

The model based on POS dependency bigrams, which achieved the best compromise between classification accuracy and linguistic interpretability in classifying the non-mediated dataset, was applied to the mediated dataset. Following the MuPDAR[F] approach (Gries & Deshors 2020), the aim of this second set of analyses is that of assessing whether and how the two datasets of native and interpreted texts differ from one another in terms of formality-related features.

Table 7: Prediction accuracy of interpreted data

	Source language			
Prediction	fr	it	pl	Total
Impromptu: correct	4 (20%)	2 (20%)	6 (30%)	12 (24%)
Read out: correct	7 (35%)	3 (30%)	7 (35%)	17 (34%)
Read out: erroneous	6 (30%)	3 (30%)	4 (20%)	13 (26%)
Impromptu: erroneous	3 (15%)	2 (20%)	3 (15%)	8 (16%)
Total	20 (100%)	10 (100%)	20 (100%)	50 (100%)

Table 7 reports data on the model's prediction accuracy in the interpreted dataset, expressed as the number of interpreted texts that were identified correctly and erroneously as deriving from an impromptu or read out source text (henceforth called "impromptu" and "read out" texts for brevity). Since the baseline is constituted by predictions on non-mediated texts, the higher the accuracy, the more mediated texts can be hypothesized to be similar to their non-mediated

counterpart. Conversely, the higher the degree of deviation, the more mediated texts can be seen as different from non-mediated ones in terms of formality.

Impromptu interpreted texts get the lowest percentage of correct predictions (24%), and the highest percentage of incorrect ones (26%), pointing to the fact that they are most often predicted as being read out. Only rarely does the opposite scenario occur, i.e. that the model predicts a read out interpreted text as being impromptu (16% of cases). The picture that emerges is thus one where mediated texts are generally predicted as being read out, irrespective of the actual mode of delivery of their source text.

If, as we hypothesized, and as the human evaluation seems to confirm, the impromptu vs. read out distinction reflects a distinction in terms of formality, the analysis suggests that 1) mediated texts tend to differ from their non-mediated counterparts in terms of levels of formality, and 2) deviations occur in the direction of mediated texts being more formal than non-mediated ones.

In order to assess which features contributed to the erroneous predictions, we built another model that focused solely on the erroneous predictions. The response variable was the gravity of the prediction error – how far the prediction was from being correct – and the predicting variables were those identified earlier as contributing to distinguish the impromptu from the read out texts. The contribution of the different variables was measured in terms of the permutation-based variable importance values reported as part of the model. Figure 2 includes only those features with positive values in permutation, i.e. those which contributed positively to modelling. In other words, Figure 2 displays which features tend to be used by interpreters differently with respect to native speakers, leading to different levels of formality in interpreted texts.[5]

The most important dependency bigram which distinguishes correctly and erroneously labelled texts is constituted by pronominal subjects in a preverbal position. As can be seen in the four leftmost elements of Figure 3, these bigrams are clearly more frequent in the correctly labelled impromptu texts than in the other predicted data. When compared with the reference data, i.e. impromptu and read out non-interpreted speeches, the distribution follows the reference data, suggesting that the interpreted impromptu texts predicted erroneously as read out are indeed more formal in this respect than other impromptu texts. Interestingly, the variable behavior is not bidirectional, as the read out texts that have been predicted erroneously as impromptu pair with the correctly predicted read out texts.

---

[5]Another way of interpreting the results of this analysis would be that features found in the final model but not here do not behave differently in erroneously labelled and correctly labelled texts.

POS dependency bigrams
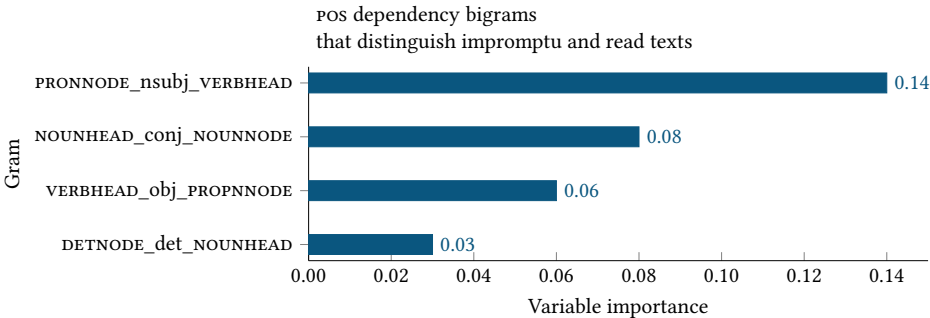that distinguish impromptu and read texts

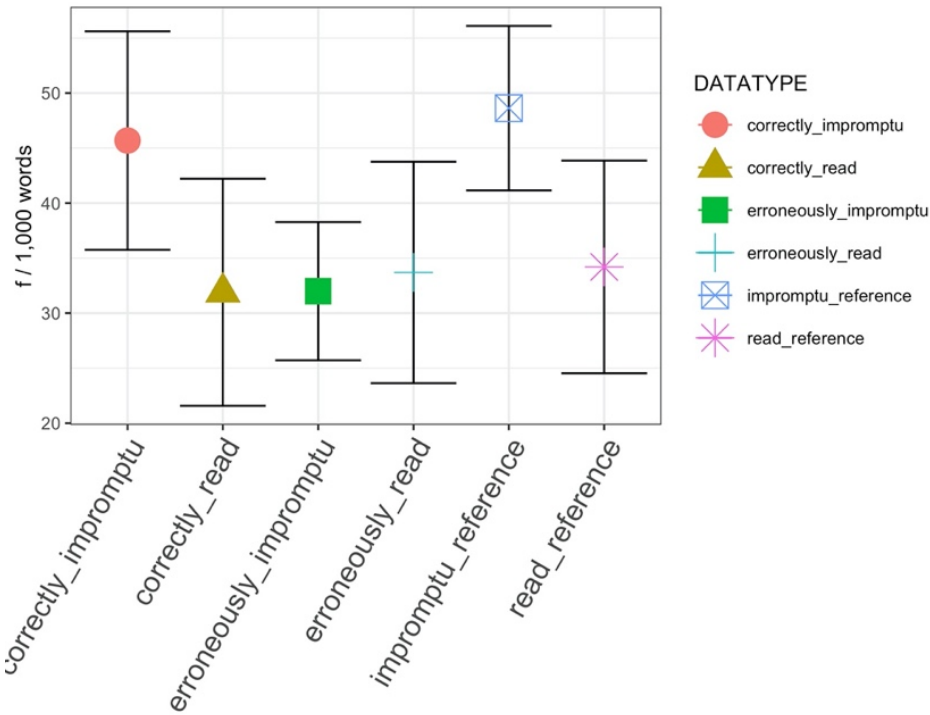Figure 2: Importance of variables that contribute to erroneous predictions

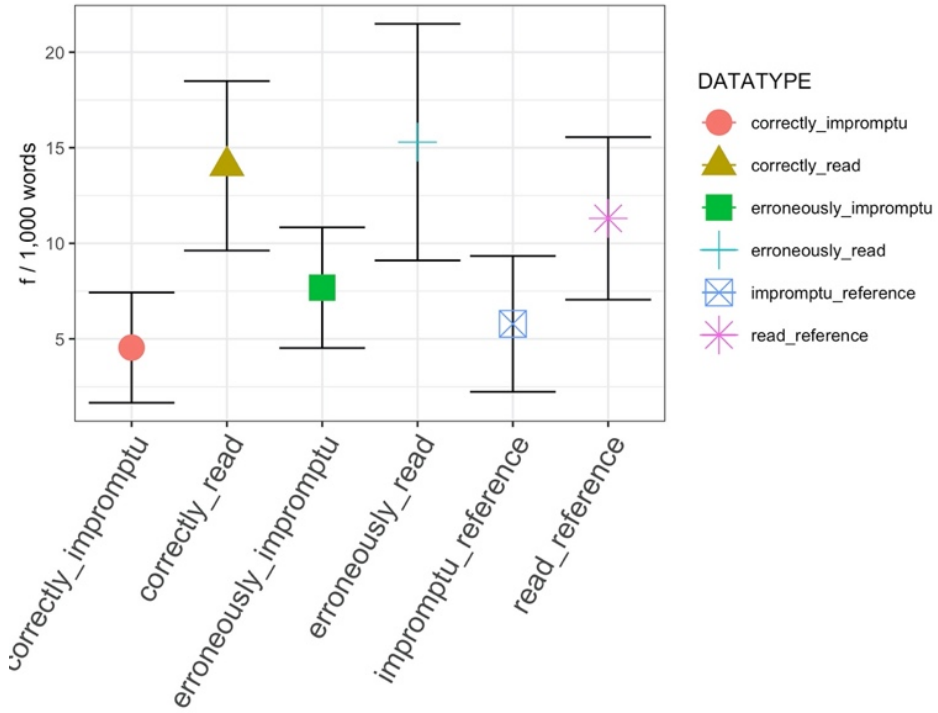Figure 3: Normalized frequencies of PRONNODE_nsubj_VERBHEAD dependency bigram

Figure 4: Normalized frequencies of NOUNHEAD_conj_NOUNNODE dependency bigram

The dependency bigram scoring second in terms of variable importance reflects the use of coordinated noun phrases. As indicated in Figure 4, the structure is more frequent in the correctly labelled read out texts than in the correctly labelled impromptu texts, and this tendency reflects the pattern in the reference data. Impromptu texts that are labelled erroneously as being read out behave similarly to the correctly labelled read out texts, while texts labelled erroneously as impromptu are closer in this regard to the texts labelled correctly as impromptu. The effect of this variable is bidirectional, as it distinguishes both texts with increased formality (predicted erroneously as read out) and those with decreased formality (predicted erroneously as impromptu).

Dependency bigrams featuring proper nouns as postverbal objects are the third most important distinguishing feature for the erroneously labelled texts. They occur more frequently in the correctly labelled read out texts than in the correctly labelled impromptu texts, and the impromptu texts labelled erroneously
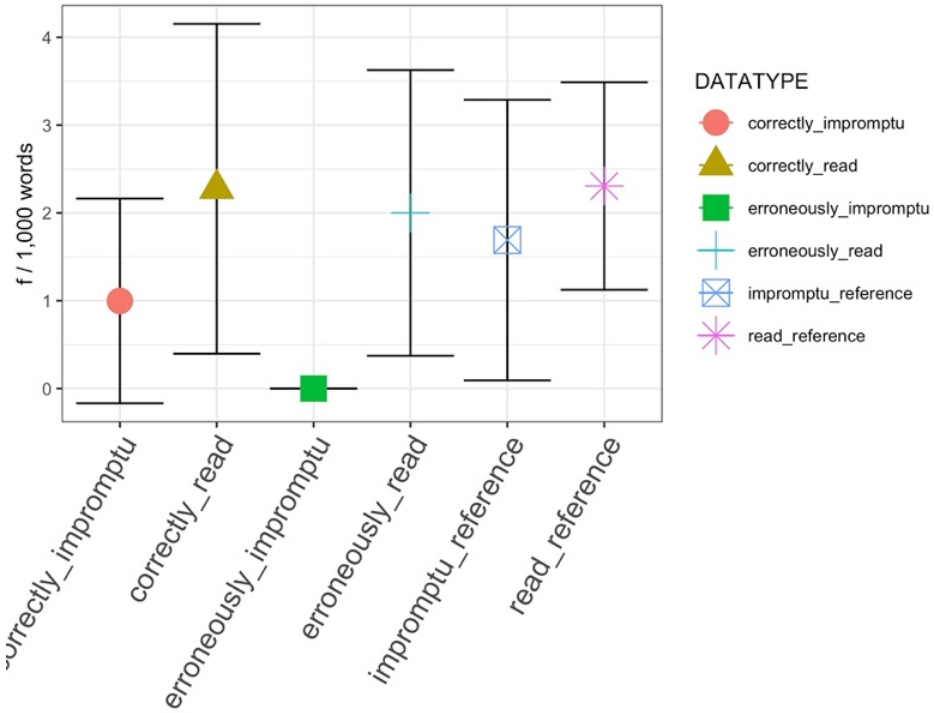
Figure 5: Normalized frequencies of VERBHEAD_obj_PROPNNODE dependency bigram

as read out are grouped together with the actual read out texts (see Figure 5). This grouping also reflects the reference data, where these bigrams are more frequent in the read out texts than in the impromptu ones. It should be noted, however, that the feature is relatively rare (only 1.5 / 1,000 words on average), and it does not occur a single time in the texts labelled erroneously as impromptu.

The fourth most important dependency bigram reflects the use of prenominal determiners. Such determiners are relatively more common in the correctly labelled read out texts than in the correctly labelled impromptu texts (see Figure 6). This tendency reflects that of the reference data, even though determiners seem to be overall more frequent in the interpreted data than the reference data. In this case, the erroneously labelled texts actually behave relatively similarly to the correctly labelled ones, but the impromptu texts labelled erroneously as read out fall in between the two correctly labelled datasets. This might be taken to suggest that they are in this respect more formal than the other impromptu texts.
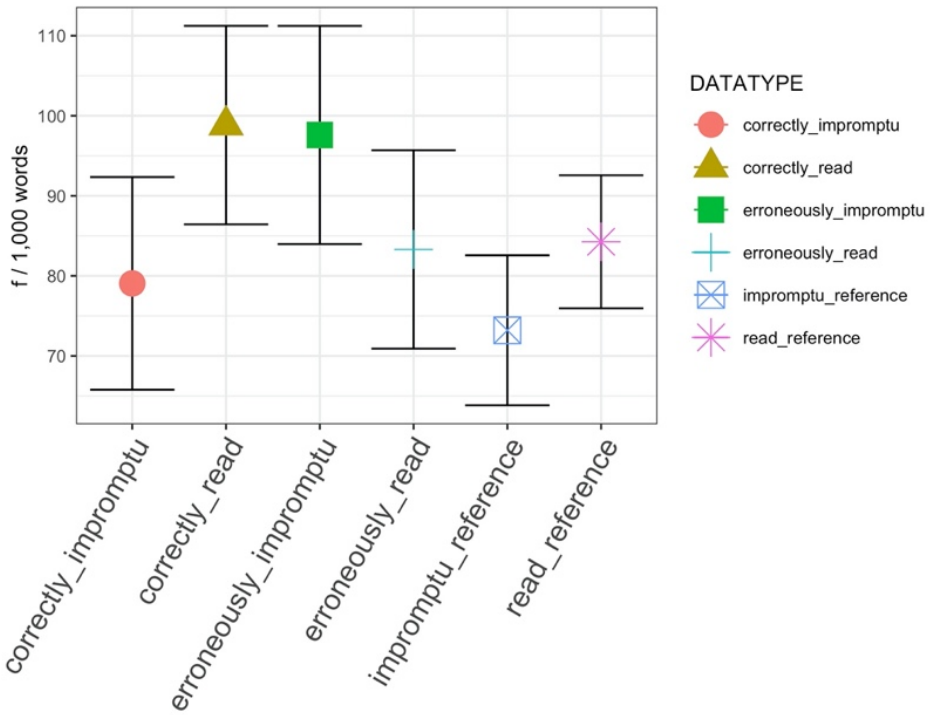
Figure 6: Normalized frequencies of ᴅᴇᴛɴᴏᴅᴇ_det_ɴᴏᴜɴʜᴇᴀᴅ dependency bigram

5 Discussion and conclusion

Formality is oftentimes referred to in comparisons across registers, genres or varieties as an explanation of the differences in linguistic features observed between them. However, so far it has rarely been the focal point of corpus-based linguistic investigations in general, and in Translation Studies in particular. As pointed out in §2, the link between formality and the features that attest to it is usually established indirectly, as formality differences are not subject to independent evaluation. In this paper we have made an attempt to fill this gap by triangulating human judgements and specific linguistic features derived bottom-up from a corpus. On the basis of the human-validated dataset of formality features, we used a corpus-based approach to examine formality differences of interpreted and non-interpreted texts.

We started by examining read out and impromptu speeches delivered at the European Parliament by native English speakers, and obtained a list of nine linguistic features contributing to text formality or informality. In an experiment involving human judgements, we observed that the distinction between read out and impromptu speeches is associated with a difference between more formal vs. more informal texts. On the basis of this evidence, we used the nine features in a model that classified interpreted texts as read out vs. impromptu (more vs. less formal). This analysis showed that interpreted texts were generally predicted as being read out, irrespective of the actual mode of delivery of their source text, pointing to a higher level of formality. Overuse of some of the features, however, pointed in the opposite direction, i.e. to informality in interpreted texts, even in cases where source texts were read out. In search of potential explanations for such results, we looked into the linguistic features that contributed to the most erroneous predictions of interpreted texts, or in other words the over- or under-represented features in interpreted texts that increased or decreased the level of formality.

One of these features involves coordinated noun phrases, which occur more frequently in the read out texts examined here, and are overrepresented in some interpretations of impromptu texts and underrepresented in some interpretations of read out texts. Nouns, in general, are typically more frequent in formal texts (Heylighen & Dewaele 1999), as formal settings usually require clarity and precision, and nouns (binomials in particular) are more likely to increase precision than, e.g., context-dependent pronouns. Yet, nouns are cognitively more demanding than pronouns, as lexical access to content words is in general slower than access to function words (Segalowitz & Lane 2000), and interpreters need to carefully manage their cognitive load, which might be decisive in this context.

Looking at it from another angle, the outcomes reported here also tap into issues long investigated by translation and interpreting scholars. The greater use of nouns instead of pronouns might hint at interpreters' explicitating meaning (Blum-Kulka 1986). Hence, the use of coordinated noun phrases, here identified as a feature of formality, might in the case of interpreters be associated to factors like cognitive load and the need to disambiguate meaning.

Postverbal proper nouns acting as objects constitute another feature that was more frequently found in read out (formal) texts, and contributed to erroneous formality classifications of interpreted texts. A large proportion of the actual expressions hidden behind these dependency tags refer to the act of thanking a specific person. In the context of the European Parliament, these formulae are used when a speaker is thanking another MEP or thanking the President for giving them the floor. Expressing thanks is a recurring act in this context and, as

also pointed out by the respondents of our survey, formulae increase the level of formality of the text.

Frequencies of use of pronominal subjects also led to erroneous classifications of interpreted text. Both pronouns and verbs are typically more frequently used in informal texts, with pronouns being deictic words referring to immediate context (Heylighen & Dewaele 1999) and associated with personal involvement. It is worth noting that interpreted texts, even though produced simultaneously in the same setting as their source texts, are a product of mediation and transferring the message of the original speaker. It is plausible that both personal involvement and immediacy of context diminish in language mediation, potentially leading to lower frequencies of pronominal subjects in a preverbal position in texts that otherwise bear more traits of informality.

Before concluding, a few limitations of the research design and method should be highlighted. First, the small size of the sample cannot be overlooked. This was mainly justified by the labour intensiveness of transcribing speeches, as well as the need to have part of them annotated by human subjects. Replication studies are therefore in order to test the results obtained here, based on larger and/or more varied datasets (e.g. in terms of text types), and ideally involving a higher number of respondents, e.g. by adopting crowdsourcing methods. While the statistical methods were selected with these limitations in mind, it is likely that richer featuresets (e.g. word trigrams) would have fared relatively better with larger datasets. On the other hand, the advantage of simpler, and arguably more abstract, featuresets such as POS dependency bigrams is that they make study designs like this feasible. Second, it should be noticed that the use of POS dependency bigrams, though reaching a satisfactory level of classification accuracy, limits the scope of the investigation to syntactic (and partly lexical) phenomena only, thus excluding features pertaining in the level of discourse, which were mentioned by respondents as being equally important as lexis in determining formality. The third and last note of caution concerns the use of a dependency parser to extract model features. Parsers are usually trained on written data, while in this case we applied them to spoken data: further studies could investigate the impact of parser accuracy in study setups like the one adopted here.

In terms of applications, we think of interpreter training as the field on which our results have a more direct bearing. As shifts of formality might have an impact on the perception of the speech, it is vital that both interpreters and interpreter trainers are sensitized to this issue and the list of features associated with formality differences in English could hopefully help in the development of adequate training aids. Hopefully, the approach demonstrated in this paper might also be instrumental in the development of interpreter training aids targeted at

sensitizing future interpreters to formality shifts in genres other than parliamentary debates.

Appendix A

A.1 Editing guidelines for spoken texts, and text selection criteria

- Eliminate DISFLUENCIES (e.g. "and **t-** tremendous concern"), EMPTY and FILLED PAUSES (e.g. "we have to **ehm** protect").

- Eliminate REPETITIONS, but only when these are in the context of other disfluencies (e.g. "procedure in this House, ehm because **we have ma-** we have been able to make significant improvements").

- Keep "Thank you President" at the beginning.

- Add punctuation, especially commas, especially in the EPIC texts, where punctuation is not present (e.g. "And we have to ask why do they do it, and" => "And we have to ask: why do they do it? And")

- Select texts with around 150 words or more. Where needed, shorten texts to make them no longer than 200 words (for gold standard/train data), and 250 words (for test set).

A.2 Example of an edited text

Transcript text	Text after edits
thank you very thank you very much ehm President. can say that where I come from in Northern Ireland we have a very vibrant poultry industry. and t- tremendous concern has been expressed to me by that industry. and what has happened while is unfortunate what has happened in Asia I think we have to ehm protect ehm the European market because it's an extremely ehm large market ehm for the poultry industry. I am concerned about the length of time it took the authorities in Asia in letting us know wha- that the the the outbreak had taken place.	Thank you very much President. I can say that where I come from in Northern Ireland we have a very vibrant poultry industry. And tremendous concern has been expressed to me by that industry. What has happened in Asia, I think we have to protect the European market because it's an extremely large market for the poultry industry. I am concerned about the length of time it took the authorities in Asia in letting us know that the outbreak had taken place.

Appendix B Text of questionnaire

Dear Participant!
Language we encounter in our everyday lives varies in many ways - spoken language diverges from written language, different dialects differ from each other, language in the school text books is different from legal documents and so on.

One of the ways different uses of language diverge from each other is **formality** - certain texts seem more formal than others. In this study, we are interested in formality and would like to ask for your help in understanding it better. In what follows, you will be shown ten pairs of texts. We would like you to **quickly read the texts and simply indicate which of the texts in each pair you find more formal**. All the texts come from the European Parliament discussions. We encourage you to follow your first instinct in the decision making.

The whole questionnaire should take about 10 minutes. After the **ten questions**, we will ask whether there was something specific that governed your decision-making. You will also be asked a couple of very basic questions on your language background. Overall, the whole questionnaire is anonymous and nei-

ther we nor anyone else have access to any personal information of the participants.

References

Abu Sheikha, Fadi & Diana Inkpen. 2010. Automatic classification of documents by formality. In *Proceedings of the 6th International Conference of Natural Language Processing and Knowledge Engineering*, 1–5. Beijing: NLPKE-2010. DOI: 10.1109/NLPKE.2010.5587767.

Andrén, Mats, Johan M. Sanne & Per Linell. 2010. Striking the balance between formality and informality in safety-critical communication: Train traffic control calls. *Journal of Pragmatics* 42(1). 220–241. DOI: 10.1016/j.pragma.2009.05. 022.

Atkinson, J. Maxwell. 1982. Understanding formality: The categorization and production of "formal" interaction. *British Journal of Sociology* 33(1). 86–117. DOI: 10.2307/589338.

Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 233–250. Amsterdam: John Benjamins.

Baroni, Marco & Silvia Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21(3). 259–274. DOI: 10.1093/llc/fqi039.

Bartłomiejczyk, Magdalena. 2016. *Face threats in interpreting: A pragmatic study of plenary debates in the European Parliament.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Biber, Douglas. 1988. *Variation across speech and writing.* Cambridge: Cambridge University Press.

Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison.* Cambridge: Cambridge University Press.

Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8(1). 9–37. DOI: 10.1515/cllt-2012-0002.

Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation. In Juliane House & Shoshana Blum-Kulka (eds.), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*, vol. 17, 17–36. Tübingen: Gunter Narr Verlag.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32.

Conrad, Susan & Douglas Biber (eds.). 2001. *Variation in English: Multidimensional studies.* Essex: Pearson.

De Sutter, Gert, Isabelle Delaere & Koen Plevoets. 2012. Lexical lectometry in corpus-based translation studies. In Michael P. Oakes & Meng Ji (eds.), vol. 51 (Studies in Corpus Linguistics 51), 325–345. DOI: 10.1075/scl.51.13sut.

Deshors, Sandra C. & Stefan Th. Gries. 2016. Profiling verb complementation constructions across New Englishes. *International Journal of Corpus Linguistics* 21(2). 192–218. DOI: 10.1075/ijcl.21.2.03des.

Fang, Alex Chengyu & Jing Cao. 2009. Adjective density as a text formality characteristic for automatic text classification: A study based on the British National Corpus. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, 130–139. Hong Kong: City University of Hong Kong. https://www.aclweb.org/anthology/Y09-1015.

Ferraresi, Adriano & Silvia Bernardini. 2019. Building EPTIC: A many-sided, multi-purpose corpus of EU parliament proceedings. In Irene Doval & M. Teresa Sánchez Nieto (eds.), *Parallel corpora for contrastive and translation studies: New resources and applications*, 123–139. Amsterdam/Philadelphia: John Benjamins.

Graesser, Arthur C., Danielle S. McNamara, Zhiqang Cai, Mark Conley, Haiying Li & James Pennebaker. 2014. Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal* 115(2). 210–229. DOI: 10.1086/678293.

Gries, Stefan Th. & Allison S. Adelman. 2014. Subject realization in Japanese conversation by native and non-native speakers: Exemplifying a new paradigm for learner corpus research. In Jesús Romero-Trillo (ed.), *Yearbook of corpus linguistics and pragmatics 2014: New empirical and theoretical paradigms*, 35–54. New York: Springer.

Gries, Stefan Th. & Sandra C. Deshors. 2014. Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9(1). 109–136. DOI: 10.3366/cor.2014.0053.

Gries, Stefan Th. & Sandra C. Deshors. 2015. EFL and/vs. ESL? A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research* 1(1). 130–159.

Gries, Stefan Th. & Sandra C. Deshors. 2020. There's more to alternations than the main diagonal of a 2×2 confusion matrix: Improvements of MuPDAR and other classificatory alternation studies. *ICAME Journal* 44(1). 69–96. DOI: 10.2478/icame-2020-0003.

Hale, Sandra. 1997. The treatment of register variation in court interpreting. *The Translator* 3(1). 39–54.

Heylighen, Francis & Jean-Marc Dewaele. 1999. *Formality of language: Definition, measurement and behavioral determinants*. Internal report. Brussels, Belgium: Center "Leo Apostel", Free University of Brussels.

Heylighen, Francis & Jean-Marc Dewaele. 2002. Variation in the contextuality of language: An empirical measure. *Foundations of Science* 7(3). 293–340. DOI: 10.1023/A:1019661126744.

Ivaska, Ilmari & Silvia Bernardini. 2020. Constrained language use in Finnish: A corpus-driven approach. *Nordic Journal of Linguistics* 43(1). 33–57. DOI: 10.1017/S0332586520000013.

Ivaska, Ilmari, Adriano Ferraresi & Silvia Bernardini. 2022. Syntactic properties of constrained English: A corpus-driven approach. In Sylviane Granger & Marie-Aude Lefer (eds.), *Extending the Scope of Corpus-Based Translation Studies*, 133–157. London: Bloomsbury. http://doi.org/10.5040/9781350143289.0013.

Jarvis, Scott. 2013a. Capturing the diversity in lexical diversity. *Language Learning* 63. 87–106.

Jarvis, Scott. 2013b. Defining and measuring lexical diversity. In Scott Jarvis & Michael Daller (eds.), *Vocabulary knowledge: Human ratings and automated measures* (Studies in Bilingualism), 13–44. Amsterdam: John Benjamins. DOI: 10.1075/sibil.47.03ch1.

Jarvis, Scott. 2017. Grounding lexical diversity in human judgments. *Language Testing* 34(4). 537–553.

Kajzer-Wietrzny, Marta. 2012. *Interpreting universals and interpreting style*. Poznań: Uniwersytet im. Adama Mickiewicza w Poznaniu. (Doctoral dissertation). https://hdl.handle.net/10593/2425.

Kanerva, Jenna, Filip Ginter, Niko Miekka, Akseli Leino & Tapio Salakoski. 2018. Turku neural parser pipeline: An end-to-end system for the CoNLL 2018 shared task. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to universal dependencies*, 133–142. Brussels, Belgium: Association for Computational Linguistics.

Koppel, Moshe & Noam Ordan. 2011. Translationese and its dialects. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1318–1326. Portland, Oregon: Association for Computational Linguistics. http://www.aclweb.org/anthology/P11-1132.

Kruger, Haidee & Gert De Sutter. 2018. Alternations in contact and non-contact varieties: Reconceptualising that-omission in translated and non-translated English using the MuPDAR approach. *Translation, Cognition & Behavior* 1(2). 251–290. DOI: https://doi.org/10.1075/tcb.00011.kru.

Kruger, Haidee & Bertus van Rooy. 2016. Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide* 37(1). 26–57. DOI: 10.1075/eww.37.1.02kru.

Kruger, Haidee & Bertus van Rooy. 2018. Register variation in written contact varieties of English: A multidimensional analysis. *English World-Wide* 39(2). 214–242. DOI: 10.1075/eww.00011.kru.

Kursa, Miron B. & Witold R. Rudnicki. 2010. Feature selection with the Boruta package. *Journal of Statistical Software* 36(11). 1–13. DOI: 10.18637/jss.v036.i11.

Li, Haiying, Arthur C. Graesser, Mark Conley, Zhiqiang Cai, Phillip I. Pavlik Jr & James W. Pennebaker. 2016. A new measure of text formality: An analysis of discourse of Mao Zedong. *Discourse Processes* 53(3). 205–232. DOI: 10.1080/0163853X.2015.1010191.

Mauranen, Anna. 2008. Universal tendencies in translation. In Gunilla M. Anderman & Margaret Rogers (eds.), *Incorporating corpora: The linguist and the translator*, 32–48. Bristol, Blue Ridge Summit: Multilingual Matters.

Olohan, Maeve. 2003. How frequent are the contractions? A study of contracted forms in the Translational English Corpus. *Target* 15(1). 59–89. DOI: 10.1075/target.15.1.04olo.

Olohan, Maeve. 2004. *Introducing corpora in translation studies*. London; New York: Routledge.

Popescu, Marius. 2011. Studying translationese at the character level. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, 634–639. Hissar, Bulgaria: Association for Computational Linguistics. https://www.aclweb.org/anthology/R11-1091.

Pym, Anthony. 2005. Explaining explicitation. In Krisztina Károly & Ágota Fóris (eds.), *New trends in translation studies. In honour of Kinga Klaudy*, 29–34. Budapest: Akadémiai Kiadó.

Rabinovich, Ella, Sergiu Nisioi, Noam Ordan & Shuly Wintner. 2016. On the similarities between native, non-native and translated texts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1870–1881. Berlin: Association for Computational Linguistics. DOI: 10.18653/v1/P16-1176.

Sandrelli, Annalisa, Claudio Bendazzoli & Mariachiara Russo. 2010. European Parliament Interpreting Corpus (EPIC): Methodological issues and preliminary results on lexical patterns in simultaneous interpreting. *International Journal of Translation* 22. 165–203.

Segalowitz, Sidney J. & Korri C. Lane. 2000. Lexical access of function versus content words. *Brain and Language* 75(3). 376–389.

Shlesinger, Miriam. 1989. *Simultaneous interpretation as a factor in effecting shifts in the position of texts on the oral-literate continuum.* Tel Aviv University. (MA thesis). DOI: 10.13140/RG.2.2.31471.69285.

Shlesinger, Miriam & Noam Ordan. 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target* 24(1). 43–60. DOI: 10.1075/target.24.1.04shl.

Toury, Gideon. 1995. *Descriptive translation studies and beyond* (Benjamins Translation Library 4). Amsterdam: John Benjamins. DOI: 10.1075/btl.100.

Volansky, Vered, Noam Ordan & Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities* 30(1). 98–118. DOI: 10.1093/llc/fqt031.

Wright, Marvin N. & Andreas Ziegler. 2017. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77(1). 1–17. DOI: 10.18637/jss.v077.i01.

Wulff, Stefanie & Stefan Th. Gries. 2019. Particle placement in learner language: Measuring effects of context, first language, and individual variation. *Language Learning* 69(4). 873–910. DOI: 10.1111/lang.12354.

Zlatnar Moe, Marija. N.d. In Daniel Gile, Gyde Hansen & Nike K. Pokorn (eds.), *Why translation studies matters* (Benjamins Translation Library 88), 125–136. Amsterdam: Benjamins. DOI: 10.1075/btl.88.12zla.