therefore whether sentence types are a better predictor of annotation quality than text type or genre, which is often postulated to be central without consideration of alternative explanations.[1]

2 Data

For our evaluation we will use the GUM corpus (Zeldes 2016)[2], a class-sourced, richly annotated multilayer corpus containing freely available texts from four different types: news articles from Wikinews, Wikimedia interviews, travel guides from Wikivoyage and how-to guides from wikiHow (abbreviated 'whow'). Each of these sources corresponds more or less to a different communicative intent, which lends itself to different types of sentences: news articles are narrative, telling about events, often in indicative past tense; travel guides are informational, giving modals of possibility and general truths about places; how-to guides are instructional, containing many imperatives and lists of ingredients; and interviews are conversational, often containing question and answer pairs or sequences. Interviews in particular could be expected to differ from the other types, due to differences between spoken and written language. The corpus contains 54 documents, totaling just over 44,000 tokens, as outlined in Table 1.

text type	source	texts	tokens
Interviews	Wikinews	14	12,661
News	Wikinews	15	9,402
Travel guides	Wikivoyage	11	9,240
How-tos (instructional)	wikiHow	14	12,776
Total		**54**	**44,079**

Table 1: Composition of the GUM corpus.

Document structure is annotated using TEI XML labels, and each text is annotated with POS tags and lemmas, dependency and constituent syntax, entities (using a subset of categories from OntoNotes, Hovy et al. 2006), information status (the scheme in Dipper et al. 2007), coreference, Rhetorical Structure Theory (Mann & Thompson 1988), and crucially, sentence type (see below). In this paper we will be concerned with:

1. POS tags – annotated manually using the extended Penn tag set used by the TreeTagger[3] (Schmid 1994)
2. Manually corrected Stanford Typed Dependencies (de Marneffe & Manning 2013)
3. Coreference annotation, including pronominal anaphora, lexical coreference and appositions (but not bridging, which is also annotated in the corpus).

Finally, the sentence type annotation layer supplies a kind of rough speech act or sentence mood, using an extended form of the SPAAC annotation scheme (Leech et al. 2003). The sentence types distinguished are given in Table 2.

tag	type	example
q	polar yes/no question	*Did she see it?*
wh	WH question	*What did you see?*
decl	declarative (indicative)	*He was there.*
imp	imperative	*Do it!*
sub	subjunctive (incl. modals)	*I could go*
inf	infinitival	*How to Dance*
ger	gerund-headed clause	*Finding Nemo*
intj	interjection	*Hello!*
frag	fragment	*The End.*
other	other predication or combination	*Nice, that!'* Or: *'I've had it, go!'* (decl+imp)

Table 2: Sentence type annotation in GUM.

Given genre metadata and sentence type annotations in the corpus, we would like to know which is a better predictor of errors on each layer.[4]

Our analyses of each data type will be addressed in separate experiments, similar in general configuration but adapted to the needs of the data type: POS tags in Section 4, dependencies in Section 5, and coreference in Section 6.

[1] An anonymous reviewer has pointed out that many other covariates of genre could be subjected to a similar treatment, in the vein of Biber's multidimensional analysis (see Biber 2009 for an overview), such as tense and other grammatical features. We agree completely: we are only beginning to understand the components of genre variation and how it interacts with annotation quality.

[2] The data is freely available under a CC license from `http://corpling.uis.georgetown.edu/gum` . We would like to thank the annotators, a current list of which is found at the same Web site.

[3] The tag set used by TreeTagger distinguishes forms of *be* (VB, 3rd person present VBZ,..) from *have* (VH, VHZ, ...) and other verbs (VV, VVZ, ...), as well as several punctuation tags and a special tag for *that* as a complementizer (IN/that). GUM also contains a second POS layer using the CLAWS5 tags (Garside & Smith 1997), which will not be evaluated here.

[4] An anonymous reviewer has asked about the decision to include the *sub* type as distinct from *decl*: this type was already in the existing annotation of GUM and was not added for this study. However modality is expressed syntactically e.g. via auxiliaries, ultimately influencing sentence structure, and semantic influence on humans should not be ruled out either.