Seminar 2

Recall & precision

"Two main parameters of retrieval effectiveness have been used over the years, defined as the proportion of relevant materials retrieved, or recall (R), and the proportion of retrieved materials that are relevant, or precision (Salton: 1989, 248); Applied to the passive, this gives the following formulae:

Precision rate: <u>number of automatically retrieved central passives</u>

Number of automatically retrieved passive tags

Recall rate: <u>number of automatically retrieved central passives</u>

Number of manually retrieved central passives

The precision rate is the proportion of retrieved forms, that is, Aux (pass) tags, that are relevant, that is, qualify as central passives. The recall rate, on the other hand, is the proportion of the relevant materials, that is, the manually retrieved central passives, that are retrieved automatically. To calculate these formulae, one needs to edit the set of automatically retrieved passives, selecting only those which fall into the category of central passives; This set of items will subsequently be referred to as the;post-edited set; that is, the set of all those automatically retrieved passives which are central passives. (Granger 1997: 370)

We can conclude from this that in the case of the passive, a fully automatic approach may conceal some relevant contrasts, but that an automatic approach combined with a manual editing stage is a reliable method of analysis (Granger 1997: 371).

In papers: Biber Tagger precision and recall rates of over 90%

References

Granger, S. (1997). Automated Retrieval of Passives from Native and Learner Corpora. Journal of English Linguistics, 25(4), 365–374. doi:10.1177/007542429702500410

Salton, Gerard. 1989. Automatic Text Processing. Reading, MA: Addison-Wesley.

Paper Cat: Анализируемые атрибуты (маркеры) стиля

Marker	Expert 1	Expert 2	Paper Cat
Abstract semantic verbs	81	81	45
Archaisms	0	0	0
Complex conjunctions	3	3	2
Desemantisized verbs	38	38	38
Future	18	16	18
IPronoun	0	0	0
Intensifying adverbs	2	2	3
Logic connectors	8	8	4
Noun	702	692	744
Noun with abstract suffix	202	134	197
OrSuffix	8	8	0
PassiveVoice	35	26	29
Past	8	4	8
Prepositive attributes	218	216	304
Postpositive attributes	8 + 131	6	130
Present	106	101	117
ThatThose	12	12	0
WePronoun	20	20	21
YouHeShe	0	0	0