

Socio-Cultural Differences in the Rhetorical Structure of Scientific Abstracts: Main Findings

The reported study has analyzed the flow of rhetorical moves in a corpus of texts in the domain of Physics: abstracts of research articles written by the native speakers of English (NS-EN) and the native speakers of Russian in English (RU-EN) and in Russian (RU-RU). These groups of writers were chosen as representatives of two distinct established academic cultures as per Galtung (1981), the Saxonic (the native speakers of English) and the Teutonic (the native speakers of Russian).

The aim was to find answers to two main questions:

1. Are there differences in the rhetorical structure (the flow of moves) between abstracts intended for members of own academic community written by scholars with different cultural backgrounds?
2. Can writers completely adapt their rhetoric to foreign conventions when writing for target audiences belonging to different academic communities in a non-native language or do their native rhetoric conventions still dominate in such discourse?

1 Annotation Study

The conducted annotation study included the comparison of three annotation schemes for rhetorical moves. After two existing schemes had proved to be inappropriate for our goals, a new annotation scheme called ASSR (Annotation Scheme with Specific and Referential categories) was developed together with Dr. Magdalena Wolska. It is based on a rhetorically important distinction between strategies of direct reporting and referencing to the full text of the paper. Detailed definitions and examples of the proposed ASSR categories can be found in Annotation Guidelines (Annotation_Guidelines_abstracts.pdf).

Four annotators annotated the corpus with ASSR categories. Two of them were Master students of Computational Linguistics, and two were Master students of Physics.

The obtained average Inter-Annotator Agreement (IAA) value was $\kappa = 0.61$, which is comparable with results of other studies focusing on the annotation of argumentative and rhetorical moves in scientific discourse. The average agreement with the final Gold Standard was 0.81.

Texts written by different groups of writers in our corpus were characterized by different IAA values (Table 1).

Table 1: Average Inter-Annotator Agreement for Different Subsets of Data

	NS-EN	RU-EN	All EN data	RU-RU	All data
Cohen’s κ	0,49	0,60	0,55	0,50	0,53
κ with partial agreement	0,57	0,67	0,62	0,59	0,61

Abstracts written in the Russian language were more difficult to annotate and had more disagreements. This may indicate that Russian academic discourse is not meant to be novice-friendly, in line with the claims of Gal-tung (1981), and does not strive for absolute clarity. The highest agreement was observed for the English-language abstracts by the native speakers of Russian. It may be related to the outsider position in a foreign academic community that requires additional effort in terms of compliance with the norms of this community in order to be accepted. As a result, these writers make their rhetorical intentions even clearer than their native-speaking peers normally do.

In general, the quality of the annotations obtained with ASSR was considered sufficient to proceed to the next step of analysis of the rhetorical structure of abstracts expressed as a sequence of moves.

2 Rhetorical Moves in the Analyzed Abstracts

2.1 Positional Distribution of Moves in an Abstract

Each sentence of an abstract has been annotated as a rhetorical move. Identical consecutive moves were merged, as the focus of interest was the flow of moves, irrespective of how syntactically elaborated each move may be. Each of these resulting moves is considered as a separate position in the rhetorical structure of an abstract.

The patterns of moves distribution across positions look different for the three groups of writers in our corpus (Figure 1, created with the TraMineR package for R (Gabadinho et al., 2010)).

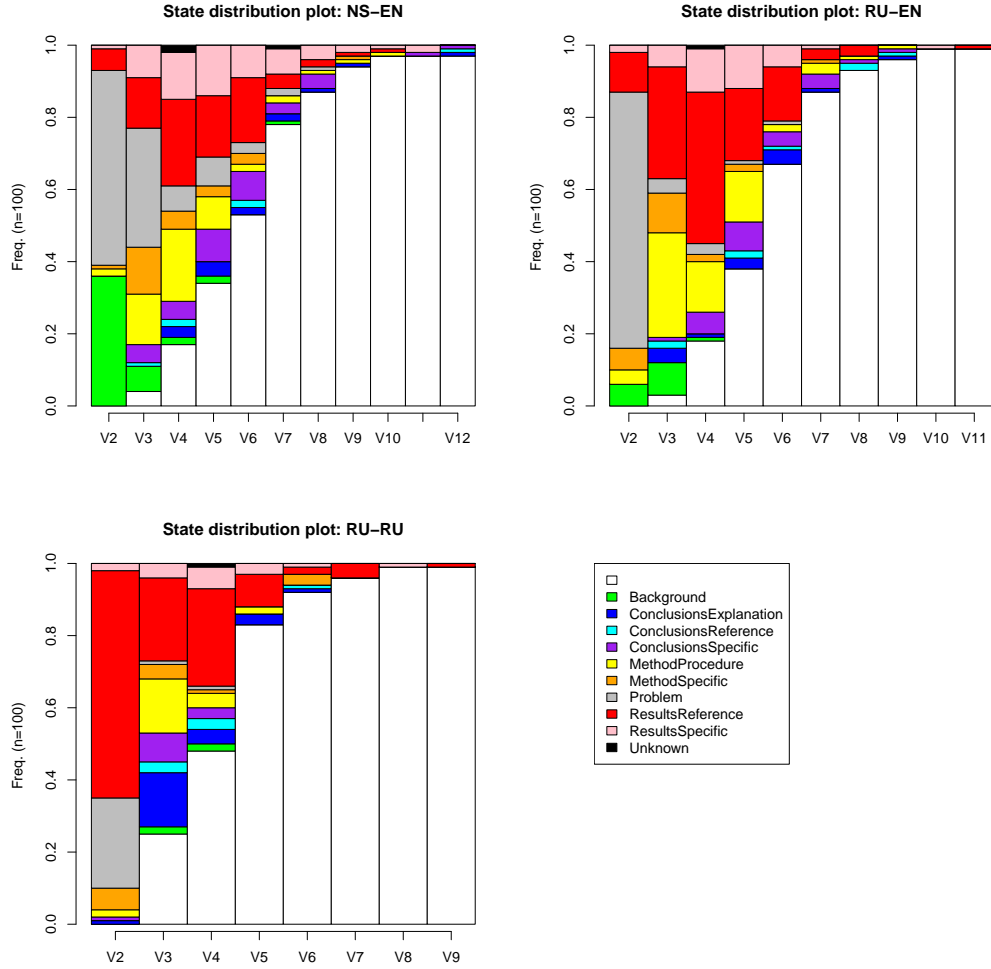


Figure 1: Distribution of moves across positions in an abstract for different groups of writers

The first move of an abstract displays most obvious differences between the groups of writers.

The native speakers of English most often start with the statement of the problem (0.54 of all moves at position 1), but also frequently use the Background move (0.36). Starting with a reference to obtained results is not typical (0.06).

The native speakers of Russian writing in English almost always start with the problem statement (0.71 of all moves at position 1). Reference to the obtained results is more frequent than for the native speakers of English

(0.11). This group of writers does sometimes use background to start an abstract, but rarely (0.06).

The group of Russian writers writing in Russian prefers to start abstracts directly with referencing the obtained results (0.63 of all moves at Position 1). This move is obviously dominating at the first position, followed with a large gap by the problem statement (0.25). Background is almost never used to start an abstract.

The middle part of an abstract also looks different for the three groups of writers.

The description of methodology is more elaborated in the abstracts written in English (both by the native speakers of English and of Russian) and is more concise in the Russian-language abstracts. Method-Procedure clearly dominates over Method-Specific for all the groups of writers, which might be a peculiar feature of the domain of Physics. Russian scholars writing in English seem to be especially careful about describing their methodology in the abstracts, probably feeling the need to invest more rhetorical effort in justifying the validity of the obtained results.

For all the abstracts, presenting the obtained results is the most important move, but it is implemented differently by different groups of writers in our corpus. Whereas for the native speakers of English Results-Reference and Results-Specific have almost equal relative frequencies, for the native speakers of Russian writing in Russian Results-Reference leaves far behind all other moves (0.65), including Results-Specific with only 0.05. The native speakers of Russian writing in English occupy an intermediate position, with 0.39 for Results-Reference and 0.13 for Results-Specific. Thus whereas the English-language abstracts employ both rhetorical strategies of direct reporting of results and reference to the results presented in the paper, Russian-language abstracts use almost exclusively the latter strategy.

The way to end an abstract looks similar for the native speakers of English and the native speakers of Russian writing in English: both prefer to end with a conclusion, the favourite move being Conclusions-Specific. Russian scholars writing in Russian often end their abstracts with Results-Reference without making any conclusion, and when they do make one, their highly preferred type is Conclusions-Explanation. The higher preference for this category shown by the native speakers of Russian writing in Russian surprisingly well matches the observation in (Galtung, 1981, p. 823) that representatives of the Teutonic intellectual style are very strong at explanation (and the Saxons are very weak at it).

Looking at the distribution of moves in an abstract, one can easily follow the trend of adaptation to foreign norms when writing for a foreign audience. The group of Russian scholars writing in English evidently moves away from

the favourite choices of their peers writing in Russian in the direction of adopting a distribution of moves typical for the native speakers of English. But the adaptation is not complete, and the writers of this group still make a number of choices that can be easily traced back to the typical distribution of moves in the abstracts written in Russian, like the more frequent use of Results-Reference and much less frequent use of Background.

In general, though the positional distribution of rhetorical moves in abstracts written by different groups of writers in our data shows some common features, it also demonstrates significant differences. The ways to start, develop and end an abstract appear to be relatively standardized within groups of writers, but not between these groups. Potentially it may lead to acceptance problems if the target audience is not limited to the members of own academic community.

2.2 Discriminative Subsequences of Rhetorical Moves

We have considered which subsequences of rhetorical moves are discriminative along the two parameters: cultural background (native speakers of English vs. native speakers of Russian) and target academic community (scholars writing in English vs. scholars writing in Russian). The results are presented in Figures 2 and 3 as TraMineR plots.

Here, the color of each bar is defined by the associated Pearson residual of the Chi-square test. For residuals less than or equal to -2 (dark red), the subsequence is significantly less frequent than expected under independence, while for residuals greater than 2 (dark blue), the subsequence is significantly more frequent (Gabadinho et al., 2010, p. 109).

Figure 2 shows five subsequences that best discriminate between the native speakers of English (NS) and the native speakers of Russian (RU). The group of the native speakers of English corresponds here to NS-EN, and the group of the native speakers of Russian includes both RU-EN and RU-RU.

In this dimension, the most relevant feature seems to be the Background move: of the five discriminative subsequences, three include this category. This rhetorical move describes the state of affairs in the field and makes the perception of subsequent information easier to the readers. It also promotes inclusion of novices and non-domain experts into the target audience.

Figure 3 shows five subsequences that best discriminate between the abstracts written in English (EN) and the ones written in Russian (RU). The first group combines NS-EN and RU-EN and the second one corresponds to RU-RU.

Here the most discriminative move is Problem. In the Russian-language

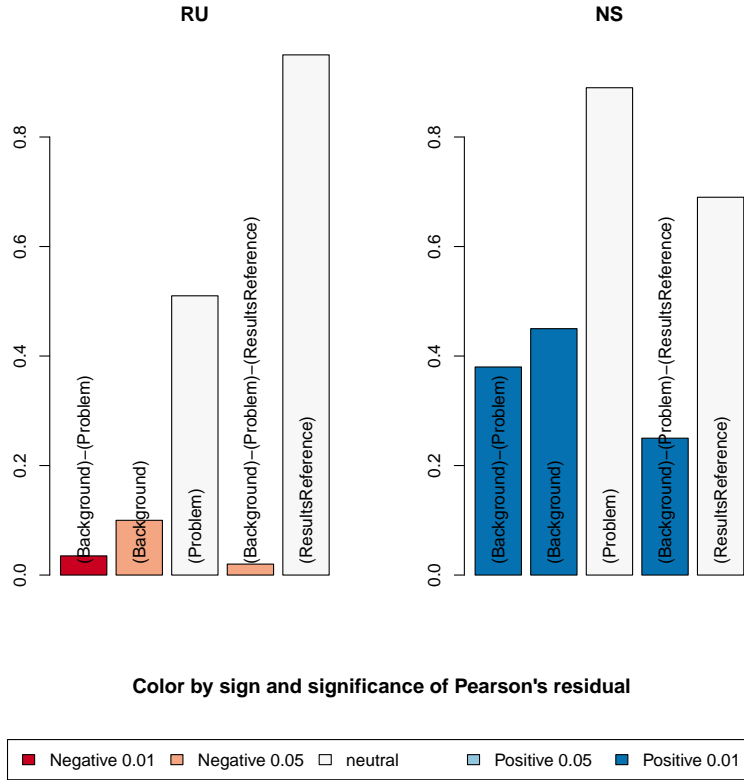


Figure 2: Discriminative subsequences of moves for writers with different cultural backgrounds

abstracts it seems to be assumed that the reader can infer the problem from the reported results and there is no need to explicitly mention it.

Thus the use of the Background move best discriminates between the writers with different cultural backgrounds, while the use of the Problem move best discriminates between abstracts for different target academic communities (in different languages).

2.3 Complete Rhetorical Structure

Looking at the full rhetorical structure of scientific abstracts one can observe some recurrent patterns that are very indicative of common trends in view of the theoretically unlimited variety of possible sequences of rhetorical moves.

Ten most frequent sequences of moves for the three groups of writers in our corpus are shown in Figure 4 (created in TraMineR).

The abstracts of the native speakers of English are characterized by a

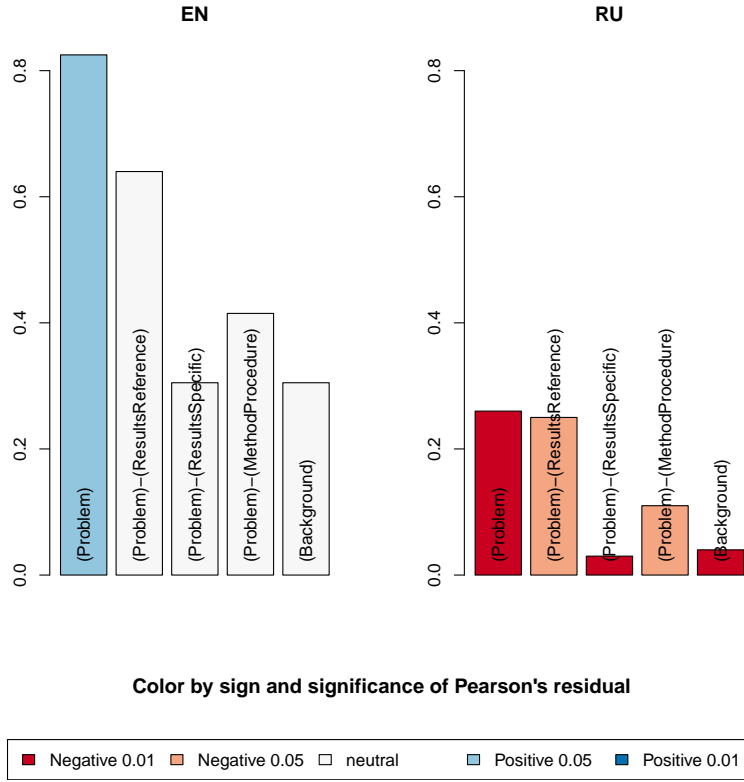


Figure 3: Discriminative subsequences of moves for abstracts written in English and in Russian

much higher diversity of patterns, and the ten most frequent sequences together constitute only 23% of data. The abstracts of the native speakers of Russian appear to be highly standardized with a small number of frequent recurring patterns. Ten most frequent sequences constitute 34% and as high as 61% of data for these groups, respectively.

For the native speakers of English, no pattern is really dominating. The three sequences that are slightly more frequent than the others are “Background – Problem” (3% of data), “Background – Problem – Results-Reference – Conclusions-Specific” (3% of data) and “Problem – Results-Reference” (3% of data). It can be noted that all these patterns very well match the CARS model by Swales (1991).

For the native speakers of Russian writing in English, four patterns are more frequent than the others, of them two are strongly dominating. The dominating patterns are “Problem – Results-Reference” (12% of data) and “Problem – Method-Procedure – Results-Reference” (7% of data), followed

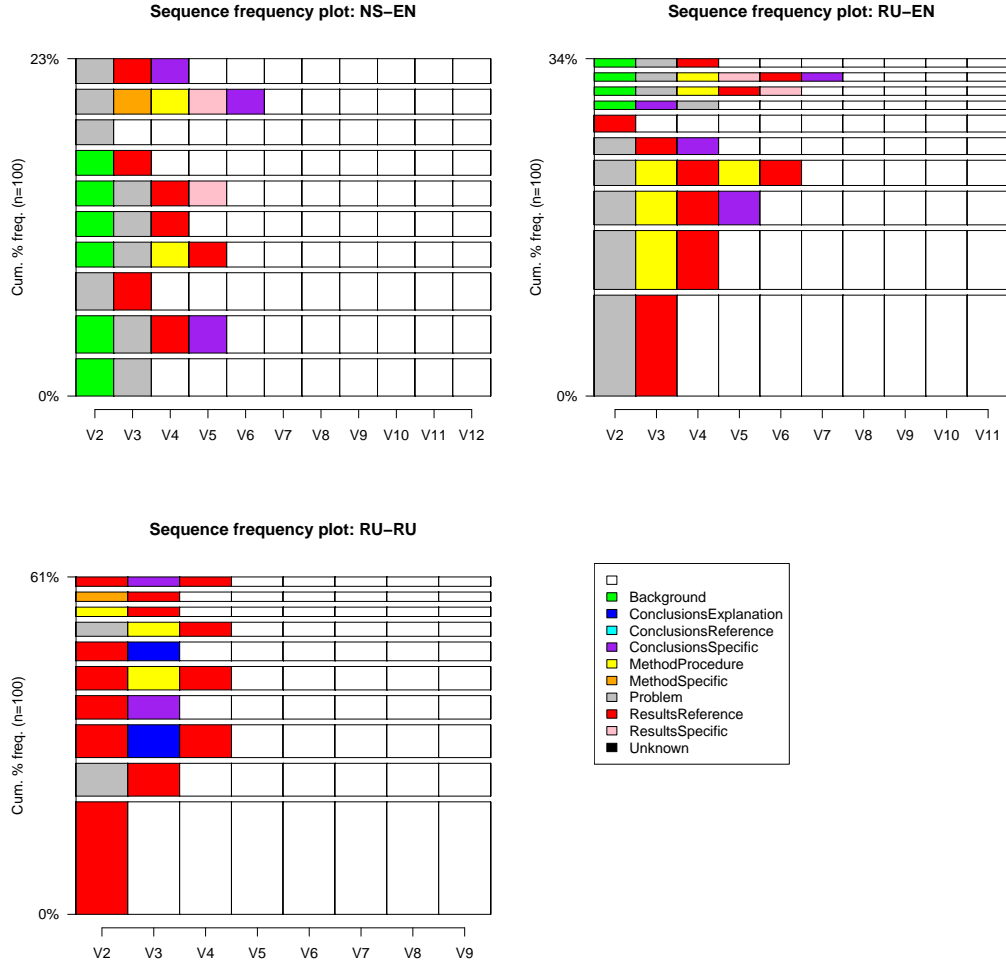


Figure 4: Frequencies of complete sequences of moves for different groups of writers

by two longer variations of the latter pattern: “Problem – Method-Procedure – Results-Reference – Conclusions-Specific” (4% of data) and “Problem – Method-Procedure – Results-Reference – Method-Procedure – Results-Reference” (3% of data).

For the Russian writers writing in Russian we can observe three most frequent patterns, of which one is very strongly dominating. As much as 24% of data for this group is covered by a single pattern “Results-Reference”, followed with a big gap by “Problem – Results-Reference” (7% of data) and the pattern not to be found in any guidelines on academic writing: “Results-Reference – Conclusions-Specific – Results-Reference” (7% of data).

Of all the most frequent sequences, only one, “Problem – Results-Reference”, is typical for all the three groups of writers and is high-ranking for all of them. This sequence seems to be the most natural pattern of an abstract in Physics. All other most frequent patterns are specific to a particular group of writers.

If we translate the categories of the classical CARS model (Swales, 1991) that describes the perfect flow of Anglo-Saxon rhetoric into ASSR categories, the ‘ideal’ scientific abstract should be a sequence of the following rhetorical moves: “Background – Problem – Results-Specific/ Results-Reference”.

In our data, only the native speakers of English really produce sequences of moves that comply with the normative requirements to academic writing in English. As soon as the pool of writers becomes culturally mixed, the actual rhetorical structures begin to differ from the ideal ones. The most frequent sequence of moves in our corpus is “Results-Reference” (9.33% of data) followed by “Problem – Results-Reference” (7.33% of data).

The domination of abstracts consisting of one single (possibly elaborated) rhetorical move of Results-Reference in our data is due to the very strong adherence of Russian writers to this pattern. This sequence seems to be the expected norm in the Russian academic community, whereas for the native speakers of English confining the whole abstract to this move, and especially starting directly with reporting results may look like an example of bad writing. Reproducing what is normal and expected in one academic community for another academic community is associated with the risk of negative evaluation and rejection by the target audience.

3 Clustering of Abstracts Based on Sequence Similarity

In order to find out if an unsupervised clustering algorithm would be able to recognize different groups of writers in our corpus based on the sequences of rhetorical moves in their abstracts we used TraMineR Ward method clustering (Gabadinho et al., 2010).

The 300 sequences were split by the algorithm into two clusters with Type 1 definitely characterizing the native Russian style of writing. Almost all (94 out of 100) abstracts in Russian were assigned to this cluster, as well as the dominating majority of English-language abstracts by the Russian writers (69 out of 100). The cluster Type 2 could be assumed to correspond to the writing style of the native speakers of English. Thus the clustering algorithm could well recognize the distinctions in the rhetoric of writing characteristic

of writers with different cultural backgrounds irrespective of the language used.

Precision, recall and F1 score values for the two classes (the native speakers of English (NS) and the native speakers of Russian (RU)) are shown in Table 2.

Table 2: Clustering Precision, Recall and F1-score

Class	Precision	Recall	F1 Score
NS	0.62	0.61	0.62
RU	0.74	0.81	0.77

Thus the distinctions previously observed were correctly captured by the unsupervised clustering algorithm that was able to differentiate two groups of writers in our corpus with a relatively high accuracy.

The most distinct group is formed by the Russian scholars writing in Russian: their sequences are almost in all cases correctly recognized as such. This also confirms that the differences in academic traditions and conventions of writing remain a reality in spite of the modern globalization processes.

The clustering results may look disappointing for the Russian scholars writing in English: however hard they may try to adopt their writing to the foreign conventions, the majority of rhetorical patterns underlying their abstracts remain essentially Russian. This may provide an answer to the question why some pieces of academic writing produced by Russian scholars look strikingly non-native and non-compliant in spite of perfectly correct English. The gate-keepers fail to explain the reason of their disappointment to the surprised authors because the difference lies at the deeper discourse levels, in the rhetorical structure and the selection of moves.

At the same time, about one-third of sequences produced by this group of writers was classified by the algorithm as belonging to the cluster of the native speakers of English, which indicates that in some cases cultural adaptation in academic writing can be quite successful.

4 Conclusions

Thus, analysis of our data has revealed significant differences in the rhetorical structure (the flow of moves) between abstracts intended for members of own academic community written by scholars with different cultural backgrounds. These differences can be well explained by the theory of academic cultures presented in Galtung (1981).

The rhetoric of English abstracts seems to be more complex and more reader-oriented. The target audience is guided through all the steps of research process starting from the state of affairs in the field to defining a problem and finding a method to solve it to obtaining results and making conclusions. Anglo-Saxon writers strive to make their abstracts accessible and easy to process for a wide audience, including non-experts. They are also much more specific in their writing than their Russian peers. This all can be seen as a manifestation of the open democratic nature of the Saxon academic culture as described by Galtung (1981).

Abstracts in Russian, on the contrary, mostly just list the results reported in the paper. Russian scholars seem to target their discourse at a limited group of peers working in the same field who can infer everything else from the provided results listing. A researcher new to the field or from an adjacent field would arguably not gain much from reading such an abstract. This rhetorical strategy can be viewed as a manifestation of the elitist and undemocratic nature of the Teutonic academic culture postulated by Galtung (1981).

The abstracts written in English by Russian scholars are of a transitive nature. In many respects they approach the norms of the native speakers of English, but in other respects still follow rhetorical conventions of the Russian academic community.

The conducted study has shown that different academic cultures remain a reality in the modern globalized world. These cultures are characterized by specific writing conventions that define the rhetoric of scientific discourse. These different approaches may have an impact on acceptance and assessment of academic writing in multicultural environments.

References

- Alexis Gabadinho, Gilbert Ritschard, Matthias Studer, and Nicolas S Müller. Mining sequence data in R with the TraMineR package: A users guide for version 1.8. 2010. URL <http://mephisto.unige.ch/traminer>.
- Johan Galtung. Structure, culture, and intellectual style: An essay comparing Saxon, Teutonic, Gallic and Nipponic approaches. *Information (International Social Science Council)*, 20(6):817–856, 1981.
- John M. Swales. *Genre Analysis. English in academic and research settings*. Cambridge University Press, 1991.