

Первый Евразийский конгресс лингвистов

Гончаров А. А., Ярошенко П. В.

Тезисы доклада

Тема: «Имплицитные логико-семантические отношения и принципы их аннотирования»

Анализ дискурса в настоящий момент является актуальным направлением в лингвистике. В рамках этой тематики рассматривается один из аспектов связности текста – логико-семантические отношения (ЛСО) между единицами с пропозициональным содержанием. ЛСО могут быть эксплицитными — выраженными при помощи языкового средства (коннектора или иного показателя), и имплицитными — не выраженными ни одним из этих средств. Так, в примере 1 содержится эксплицитное ЛСО причины (выражено при помощи *потому что*), а в примере 2 — имплицитное.

(1) *Она опоздала, потому что долго не было автобуса.*

(2) *Она опоздала, долго не было автобуса.*

Благодаря наличию показателей поиск эксплицитных ЛСО не вызывает серьезных сложностей, а определить тип ЛСО нередко можно по семантике показателя. В случае с имплицитными ЛСО все это оказывается невозможным, проблема их описания и поиска остается не решенной. Именно имплицитные ЛСО и находятся в центре внимания настоящего доклада.

Отдельной задачей в рамках изучения ЛСО является создание дискурсивно аннотированных корпусов. Принципы аннотирования корпусов разрабатываются исходя из выбранного создателями корпуса теоретического подхода к описанию структуры дискурса. В докладе анализируется специфика аннотирования имплицитных ЛСО в рамках подходов, ориентированных на анализ глобальной структуры текста (Taboada, Mann 2006), локальной структуры текста (Prasad et al. 2017), а также подхода, нацеленного на унификацию данных, размеченных в соответствии с различными теориями, и выработку единого стандарта аннотирования (Fu 2022; Zufferey, Degand 2024). Предлагается разработанный с учетом мирового опыта подход к аннотированию имплицитных ЛСО в параллельных текстах (языковые пары русский-французский, русский-английский). Объектом аннотирования является переводное соответствие, т.е. пара фрагментов оригинала и перевода, где в русскоязычном тексте показатели ЛСО отсутствуют, а в тексте на другом языке, напротив, имеются. В процессе аннотирования формируются блоки текста оригинала и перевода (их объем должен быть необходимым и достаточным для анализа ЛСО), выделяются аргументы ЛСО и показатели (при наличии), затем выявляются признаки блоков текста. Таким образом, результатом работы становится аннотированное переводное соответствие (сохраняется в базе данных).

Признаки разделены на группы в зависимости от того, к какой именно части блока текста они относятся. В докладе будут рассмотрены признаки блока текста в целом (тип ЛСО); признаки, относящиеся к одному из аргументов (синтаксическая характеристика; указание на лексические единицы, значимые для анализа ЛСО); признаки, относящиеся к обоим аргументам (типы соотношения аргументов на лексическом уровне; на уровне пунктуации); признаки переводного соответствия (например, можно указать на то, что в переводе изменена структура блока текста оригинала).

Будут представлены следующие принципы аннотирования имплицитных ЛСО: 1) необходимо выделять границы аргументов ЛСО (обеспечивается наглядность и удобство анализа); 2) признаки блоков текста должны образовывать иерархическую структуру (обеспечивается удобство работы с большим числом признаков); 3) если признак блока текста имеет лексический показатель, то этот показатель должен быть выделен (обеспечивается более высокая обоснованность решений разметчика).

Список литературы

1. Fu Y. Towards Unification of Discourse Annotation Frameworks. Muresan S., Nakov P., Villavicencio A. (eds.). Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: Association for Computational Linguistics, 2022. P. 132–142.
2. Prasad R., Webber B., Joshi A. The Penn Discourse Treebank: An Annotated Corpus of Discourse Relations. Ide N., Pustejovsky J. (eds.). Handbook of Linguistic Annotation. Dordrecht: Springer Science+Business Media, 2017. P. 1197–1217.
3. Taboada M., Mann W. C. Rhetorical Structure Theory: Looking back and moving ahead // Discourse Studies. 2006. 8(3). P. 423–459.
4. Zufferey S., Degand L. Connectives and Discourse Relations. Key Topics in Semantics and Pragmatics. Cambridge: Cambridge University Press, 2024.