

ЛИНГВИСТИЧЕСКАЯ ИДЕНТИЧНОСТЬ КОМПОЗИТА

При изучении структуры композитов найдены несоответствия в их графическом обозначении. Это свойственно для иноязычных композитов, которые составили 15% от общего количества выявленных композитов.

Заимствования функционируют в русском языке в соответствии с его нормами, что зачастую отличает их от композитов в родном языке. Слова были найдены параллельно в двух поисковых системах Yandex, Google, чтобы посредством сравнительного анализа проверить идентичность их графических оболочек. Были выявлены не только несколько вариантов написания одного композита, а также тождественные им словосочетания.

Так как в поисковых системах идентифицированы все заданные формы исследуемых композитов в контекстах, это объясняет возможность их двоякого существования в речевой системе.

Автоматизированная идентификация помогает ускорить отбор исследуемого материала, но учитываются неточности при запросе, который можно устранить указанием дополнительных параметров поиска: дат, типов файлов, стран, языков и т. д. Чем точнее запрос, тем точнее будет идентифицирована информация поиска.

На практике сбор исследовательского материала методом сплошной выборки в печатных и рукописных источниках является трудоемким процессом, уступающим автоматизированным поисковым системам в количестве затраченного времени, но позволяющим следовать меняющимся тенденциям.

На данном этапе автоматизированные поисковые системы функционируют с большим количеством погрешностей, над которыми нужно работать, но они способны обрабатывать огромные объемы информации, выявляя закономерности и обобщая полученные результаты, в считанные секунды.