

Статистическое исследование количественных показателей лексической многозначности в русском языке

Головкин Николай Вячеславович

Аспирант Ставропольского государственного университета, Ставрополь, Россия

Процесс полисемантизации, т. е. увеличения количества лексических значений слов и преобразования однозначных слов в многозначные, традиционно рассматривается как один из основных факторов развития языка, обеспечивающих обогащение его семантического состава. Наряду с заимствованием, полисемантизация может также рассматриваться как процесс наращивания объема словарного состава языка. Соответственно, логично будет предположить, что количество и массовая доля полисемантических слов в языке нарастает в определенной прогрессии, следовательно, имеет некоторые количественные показатели, характеризующие этот процесс и его интенсивность. Для их обозначения считаем необходимым задать следующую рабочую терминологию:

1. *Коэффициент степени полисемантизации* (КСП). Определим данный термин как показатель, характеризующий отношение количества многозначных слов в языке к общему количеству лексических единиц языка. Таким образом, изменение КСП характеризует динамику массовой доли полисемантических слов в языке (т. е. в диахронии), а значение КСП выражает упомянутую долю синхронически.

2. *Коэффициент интенсивности полисемантизации* (КИП) может быть определен как показатель, характеризующий интенсивность нарастания количества и массовой доли полисемантических слов в языке за определенный промежуток времени.

Для подсчета может быть применена словарная норма требуемого периода времени.

Очевидно, что КСП может быть рассчитан по формуле $\frac{QP}{QO}$, где QП – количество полисемантических слов в словаре, QO – общее количество слов в словаре; КИП – по формуле $\frac{PQP}{RQP}$, где PQП – количество полисемантических слов в более позднем издании словаря, RQP – количество полисемантических слов в более раннем издании словаря. Предпримем попытку расчета коэффициентов степени и интенсивности полисемантизации для русского языка.

В силу того, что словари разных редакций и авторов могут существенно различаться по составу лексических единиц, то для расчета нам представляется необходимым использование различных изданий одного словаря. Мы использовали при расчетах исправленные и дополненные издания словаря С.И. Ожегова – второе и девятое. Следует помнить, что отражение новых норм в исправленных изданиях словарей не всегда соответствует существующему в языке семантическому составу слова, поэтому в живой речи коэффициенты могут оказаться выше, и мы рассматриваем получаемые нами показатели как минимальные.

Отметим, что состав движущих сил развития языка, влияющих на пополнение его словарного запаса, не ограничивается лишь полисемантизацией, и одной из таких сил, находящихся отражение в словарной норме, является также и неологизация. Осуществление расчетов без принятия во внимание этого фактора может привести к смешиванию полисемантизации и неологизации, потому мы допускаем, что большинство новых слов в языке являются в момент заимствования однозначными, и не принимаем во внимание количественное пополнение словарного состава. Так, второе издание словаря содержит 52 000 слов, девятое – 57 000 слов; логично предположить, что состав словаря пополнен 5 000 однозначных слов, не относящихся к расчетам коэффициентов степени полисемантизации. Вместе с тем для полноты картины мы

рассчитываем коэффициенты применительно как к количеству слов в раннем издании словаря, так и к количеству слов в позднем издании.

Мы также не принимаем во внимание количественное пополнение семантических составов слов, уже являвшихся многозначными в предстоявшем издании словаря, поскольку этот процесс не оказывает влияния на рассчитываемые коэффициенты.

В результате расчетов нами получены данные, свидетельствующие, что количество полисемантических слов в словаре увеличилось с 8 951 до 10 086, коэффициент интенсивности полисемантизации составил 1,174 (117 %), коэффициент степени полисемантизации изменился с 0,1652 (16,5 %) до 0,1939 (19,4 %). Прогресс КСП составил 0,02875 (2,9 %). Так как число является положительным, мы имеем основания утверждать, что в языке происходит постепенное увеличение не только количества многозначных слов, но и числового выражения их массовой доли по отношению к общему количеству слов в языке.

Помимо описанных коэффициентов, возможен также расчет *локальных коэффициентов полисемантизации (ЛКП)*, т. е. КСП и КИП для слов на конкретные буквы, рассчитываемый как отношение количества полисемантических слов, начинающихся на одну букву, к общему количеству слов на данную букву. Нами были проведены расчеты ЛКСП для русского языка по данным девятого издания словаря Ожегова; в частности, они позволили установить, что массовая доля полисемантических слов на одну букву существенно варьируется в зависимости от начальной буквы слова – диапазон значений ЛКСП составил от 0,17 (17 %) для буквы Й до 0,37 (37 %) для буквы Щ. Соответственно, существуют основания для утверждения, что в русском языке можно выделить словарные группы с низкой степенью полисемантизации (< 27 %) и высокой степенью полисемантизации (> 27 %). Были получены также и данные по локальному коэффициенту интенсивности полисемантизации для каждой словарной группы; наиболее активный прирост количества многозначных слов был отмечен для буквы Д, где ЛКИП составил 1,35 (135 %), наименее активный – для буквы Й (ЛКИП = 1, т. е. прирост отсутствует).

Таким образом, мы считаем правомерным заключить, что, по полученным нами статистическим данным, в языке происходит рост количества многозначных слов, а также их массовой доли в лексиконе. Полученные данные могут способствовать прогнозированию изменений словарного состава языка, а также применяться в целях формализации ряда параметров текста при его автоматизированной обработке.

Литература

- Ожегов С.И. Словарь русского языка. Изд. 2-е, испр. и доп. М., 1950.
Ожегов С.И. Словарь русского языка. Изд. 9-е, испр. и доп. М., 1972.