Постоев И.Е. ПИН-22М

Каким образом можно выразить эмоцию с помощью текста? Как с помощью компьютера можно определить, несет ли текст какую-то эмоцию?

Текст, сам по себе, без человеческого восприятия, - просто набор символов. С его помощью человек транслирует ту или иную мысль, и каждый читатель может воспринять этот текст по-своему.

Мысль может передаваться явно, в виде информации в виде нейтрального текста без неоднозначности и эмоционального оттенка. Так происходит, например, в текстах научного или официально-делового стилей.

В силу гибкости и “естественности” естественного языка, текстом также пользуются как средством для трансляции чувств, эмоций читателю. Прочтение такого текста может спровоцировать у читателя какие-либо чувства. Это может использоваться для придания тексту скрытого смысла и/или для эмоциональной окраски. В художественных текстах это нужно, например, для создания атмосферы происходящего или раскрытия образа героя.

И если в таком контексте перефразировать первый вопрос, то, получается, необходимо определить, как формируется эмоция в тексте и какие языковые средства позволяют придать эмоциональный оттенок тексту.

В русском языке таких средств предостаточно! И в этом утверждении уже видны два из них. Использование приставки пре- в последнем слове усилило его значимость, а восклицательный знак придал ему эффект возгласа. В совокупности получилась сильная фраза-лозунг. Хотя эта же связка в предложении “У нас и своих дураков предостаточно!” несет совсем другую эмоцию за счет смысла. То есть, итоговая эмоция и то, как фраза будет интерпретирована, зависит от совокупности контекста предложения(информации), значения слов и используемых “эмоциональных” приемов.

Если обратить внимание на построение предложения, то экспрессивности может добавить пунктуация: восклицательные, вопросительные(особенно в риторических связках) знаки, многоточие(недосказанность идеи).

Если же речь идет об эмоциональной окраске слова, то данная тема раскрывается в разделе лингвистики “Лексическая стилистика”. Она выделяет богатый набор средств создания эмоционального оттенка слова.

В их числе:

* лексические повторы;
* использование слов с ярким оценочным значением(пустомеля, разгильдяй);
* многозначные слова, обычно нейтральные в основном значении, но получающие яркую эмоциональную окраску при метафорическом употреблении(о человеке: лиса, шляпа, медведь);
* слова с аффиксами субъективной оценки, передающие различные оттенки чувства(маленько, детинушка, сыночек, потихоньку);
* использование синонимов, имеющие большую экспрессивность относительно номинального слова(необщительный, грубый, черствый, неприветливый) - эпитеты;
* использование устаревших слов;
* канцеляризмы - придание официальности словам в тексте стиля, отличного от официально-делового;
* художественные приемы: антитеза, оксюморон, гипербола, сравнение и т.п;
* фоника: звукопись, звуковые повторы;
* инверсия порядка слов;

...

Выше перечислены не все способы эмоциональной окраски текста в русском языке. Но даже количество перечисленных внушающее и говорит о сложности вопроса восприятия эмоциональных оттенков текста машиной. Однако, объединяя близкие по экспрессии слова в лексические группы, можно выделить: 1) слова, выражающие положительную оценку называемых понятий, 2) слова, выражающие их отрицательную оценку. В первую группу войдут слова высокие, ласкательные, отчасти шутливые; во вторую - иронические, неодобрительные, бранные и др.

Таким образом, даже имея большой набор способов экспрессии и выражаемую ими широкую гамму чувств, эмоций, которые даже человеком могут быть не всегда восприняты однозначно, можно данную упростить задачу до определения наличия эмоциональной окрашенности(возможно, ее степень) или общего настроения текста (положительно/отрицательно). Последним занимается область NLP(Natural language processing) - сентимент-анализ.

Теперь, имея задачи и сведения о создании эмоционального оттенка в языке, предположим, как ее можно решить с помощью компьютера. Компьютер, конечно, нужен только для исполнения алгоритма и входа/выхода данных. В качестве данных может быть любой текст естественного языка(художественное произведение, сводка новостей, политический отчет и т.п).

В силу того, что средств экспрессии много и не все они так очевидны в тексте(как научить компьютер распознавать метафору?), то, допустим, на основании нескольких приемов, выделились бы критерии, например: наличие/количество эпитетов, “ярких слов”, использование аффиксов для окраски слова, использование фразеологизмов.

Видно, что для них, помимо анализируемого текста, понадобились бы наборы вспомогательных данных. Понадобился бы словарь эпитетов, “ярких слов” и фразеологизмов. Его можно получить из открытых источников. Словарь нейтральных слов и эмоционально окрашивающих их аффиксов можно было бы получить, сравнивая тексты художественного и официально-делового/научного стилей и определяя, какие для одних и тех же слов в разных стилях какие используются или не используются аффиксы. Также, было бы полезным иметь словари жаргонизмов, часто употребимых слов в официально-деловой и публицистической литературе, ведь по их присутствию можно также делать выводы о нейтральности текста (не часто встречаются тексты, где слова “надлежащий” и “чудесный” стоят рядом).

Таким образом, имея данные словари, исходный текст и компьютер, можно проверить наличие или отсутствие элементов из словарей в анализируемом тексте на естественном языке и делать вывод о наличии эмоциональной окраски текста и/или ее направленности.

Литература

1. Голуб И.Б. Стилистика русского языка <http://www.hi-edu.ru/e-books/xbook028/01/>
2. Словарь эпитетов русского языка https://gufo.me/dict/epithets