

В кн. «Информация и кибернетика». Сб. статей под ред. акад. А. И. Берга. Изд. «Сов. радио», 1967.

4. А. Н. Леснтьев, Д. Ю. Панов. Психология человека и технический прогресс. В кн. «Философские вопросы высшей нервной деятельности и психологии». Изд. АН СССР, 1963.

5. А. А. Смирнов. Проблемы психологии памяти. Изд-во «Просвещение», 1962.

6. W. R. Garner. Uncertainty and structure as psychological concepts. New York — London, Wiley, 1962.

7. E. S. Robinson, C. W. Darrow. Effect of length of lists upon memory for numbers. Amer. J. Psychol., 1924, vol. 35, pp. 235—243.

8. C. E. Shannon. A mathematical theory of communication. Bell Syst. Tech. J., vol. 27, pp. 379—423. Русск. пер. в кн. К. Шеннон. Работы по теории информации и кибернетике. Изд-во иностр. лит., 1963.

## ОБ ИНФОРМАЦИОННЫХ ИЗМЕРЕНИЯХ СПЕЦИАЛЬНЫХ ЯЗЫКОВ

*П. Б. Невельский, М. Д. Розенбаум*

В больших системах совместно работают люди-операторы и электронные вычислительные машины. Для общения человека с машиной на современном этапе развития техники вполне возможно использование не только искусственных, но и естественных языков. Как показал Д. Ю. Панов [3, 4], в процессе развития естественных языков наблюдается повышение их информативности, и взрослый развитый человек стремится использовать более информативный язык. Для машины же проще использовать менее информативный язык. Поэтому предполагается, что общение человека с машиной должно осуществляться на основе некоторого компромисса между этими двумя противоположными тенденциями. Успешное решение проблемы общения человека с машиной, очень важное для будущего развития человечества, невозможно без исследований, результаты которых позволят определить, должен ли быть найден некоторый критерий, ограничивающий информативность человеческих языков для общения с машиной, или же нужно стремиться к тому, чтобы машины делались пригодными для общения на наиболее информативных языках [3, 4].

Разрешение этой проблемы затруднено из-за целого ряда нерешенных задач. Одной из них — информационным измерениям естественного языка, точнее, специального делового (профессионального) подязыка — посвящено настоящее исследование.

Рассмотрим два информационных параметра языка, его энтропию и избыточность.

Под энтропией языка будем понимать среднюю предельную условную энтропию на одну букву незнакомого текста как меру ее неопределенности, когда известна цепочка предшествующих букв этого же текста длиной от тридцати букв и более.

Относительной энтропией будем называть отношение энтропии текста  $H$  к максимальному значению  $H_{\max}$ , которого могла бы достичь энтропия текста при тех же символах. Относительная энтропия — показатель максимального сжатия, которое в принципе можно осуществлять кодированием без изменения алфавита.

Избыточность по [6] — единица минус относительная энтропия — является мерой организации символов, мерой ограничений, накладываемых на свободу их выбора, т. е.

$$R = 1 - \frac{H}{H_{\max}}. \quad (1)$$

Если мы говорим об избыточности языка или текста, равной 0,80, или 80%, то это значит, что появление букв текста на 80% связано со связанными вероятностными ограничениями (морфологическими, синтаксическими, семантическими, логическими и др.) и что при порождении речи мы свободны в своем выборе только на 20%, которые предоставляет относительная энтропия.

Невозможно оценить энтропию языка обычными статистико-информационными методами, так как для этого нужно было бы знать вероятности распределения всех возможных, по крайней мере, тридцатибуквенных соединений, что намного превышает наши вычислительные возможности.

Преодолеть эти трудности помогли психологические опыты. В современном языкознании и математической теории связи язык рассматривается как код с вероятностными ограничениями. При этом предполагается, что по своей структуре язык близок к оптимальному коду. Это позволяет думать, что человек — носитель языка, владея вероятностными оценками относительно его элементов, пользуется этими оценками при восприятии (декодировании) текста и его порождении (кодировании сообщения). На этих предположениях строится методика угадывания неизвестного текста, впервые предложенная К. Шенноном [7], а впоследствии усовершенствованная и дополненная Р. Г. Пиотровским, А. Н. Колмогоровым и др. [5].

Задача проведенного эксперимента заключалась в сопоставлении субъективных оценок энтропии и избыточности специального текста (относящегося к очень узкому профессиональному языку), получаемых при угадывании этого текста специалистами и неспециалистами в данной области, а также специалистами, в равной мере знакомыми с подязыком текста, но в разной мере знакомыми с самим текстом.

В качестве специального делового языка был взят язык документации, проходящей и обрабатываемой в финансовом отделе одного из производств Донецкой области. Текст объемом в тысячу букв был взят из «Журнала-ордера» по затратам на производство». Этот «Журнал» представляет собой развернутый формуляр, ежедневно заполняемый цифровой информацией. Он содержит большое число рядов и колонок, наименования которых неизменны в течение нескольких лет, меняется только цифровая информация. Эти наименования и были взяты в качестве текста для угадывания:

ЮЖН Р-К ДАЛЬНИЙ Р-К ВСКРЫША ДОФ ИТОГО ЖДЦ ПЕРЕВОЗКА ПОГРУЗКА ЖДЦ ДАЛЬНИЙ УСЛУГИ ШПАЛОПРОПИТКА ПАРКОТЕЛЬНАЯ ВОДА УМЯГЧЕННАЯ ВОДА СЫРАЯ ЖД КРАНЫ ОТВАЛЬНОЕ Х-ВО ГАЗВОДА ЦРМЦ КИП ЭЦ ЭЛЕКТРОПОДСТАНЦИЯ СВЯЗЬ КИСЛОРОДНАЯ ЦКДМ РАБОТА ОСНОВ ЦЕХОВ ПАРКОТ СМУ КОМПРЕССОРА (им. пад. множ. ч. — П. Н., М. Р.) ЦКАТ ДРОБЛЕНИЕ КОНВЕЙЕР СКЛАД ЗАПАС ПО СБОРУ ЛОМА.

Этот текст, содержащий очень много сокращенных слов, высоко информативен, но почти совсем непонятен непосвященному человеку. Если этот текст «перевести» на общеизвестный русский язык, то он увеличился бы более чем вдвое.

Опыты проводились по методике [2]. Субъективная энтропия текста рассматривалась здесь как ненадежность угадывания, как ненадежность передачи информации с входа на выход, от буквы текста к ответу испытуемого. Ненадежность передачи информации — это условная энтропия входа, когда известен выход — условная энтропия текста, когда известен ответ испытуемого. Другими словами, это мера той неопределенности, которая остается в тексте после угадывания, или то количество инфор-

мации, которое испытуемый не в состоянии извлечь из текста, несмотря на знание предшествующей последовательности букв этого текста и ввиду своего прошлого опыта, т. е.

$$H_{\text{суб}} = H_y(x) = H(x, y) - H(y), \quad (2)$$

где  $H_y(x)$  — ненадежность угадывания;

$H(x, y)$  — совместная энтропия входа и выхода (диграммы «буква текста — буква в ответе испытуемого»);

$H(y)$  — энтропия выхода (угадывания)

или

$$H_{\text{суб}} = -\sum p_{ij} \log_2 p_{ij} + \sum p_j \log_2 p_j, \quad (3)$$

где  $p_{ij}$  — наблюдаемая вероятность (частота или относительная частота) совместного появления  $i$ -й буквы в тексте и  $j$ -й буквы в ответе испытуемого;

$p_j$  — наблюдаемая вероятность появления  $j$ -й буквы в ответе испытуемого.

Первую группу испытуемых из двух человек составляли специалисты со средним образованием и большим стажем работы по специальности, изо дня в день работающие с «Журналом» и вписывающие цифровую информацию в графы этого журнала, текст из которого они угадывали. Другими словами, это носители специального языка, хорошо знакомы с текстом.

Группы испытуемых	Испытуемые	Энтропия текста $H_{\text{суб}}$ , дв. ед.	Избыточность $R$ , %	Переданная информация $T(x, y)$ , дв. ед.	Ранг варианты $r$
Специалисты, хорошо знакомые с текстом	Х-ов	0,49	90,2	3,92	20
	П-ва	0,58	88,4	3,83	19
	<i>Среднее</i>	<b>0,53</b>	<b>89,4</b>	<b>3,88</b>	<b>19,5</b>
	<hr/>				
Специалисты, менее знакомые с текстом	Ш-ва	0,82	83,6	3,59	18
	З-ко	0,87	82,6	3,54	17
	Д-на	0,88	82,4	3,53	16
	Ф-ин	0,97	80,6	3,44	15
	<i>Среднее</i>	<b>0,89</b>	<b>82,2</b>	<b>3,52</b>	<b>16,5</b>
<hr/>					
Специалисты, не знакомые с текстом	И-ов	1,16	76,8	3,25	14
	Я-ко	1,18	76,4	3,23	13
	Б-ва	1,20	76,0	3,21	12
	О-юк	1,22	75,6	3,19	11
	<i>Среднее</i>	<b>1,19</b>	<b>76,2</b>	<b>3,22</b>	<b>12,5</b>
<hr/>					
Неспециалисты	М-ик	2,08	58,4	2,33	10
	В-на	2,80	44,0	1,61	9
	Г-ко	2,81	43,8	1,60	8
	О-ва	2,88	42,4	1,53	7
	М-ко	2,90	42,0	1,51	6
	О-ра	2,94	41,2	1,47	5
	С-ко	3,00	40,0	1,41	4
	К-ва	3,01	39,8	1,40	3
	Д-ко	3,16	36,8	1,25	2
	Ш-да	3,18	36,4	1,23	1
	<i>Среднее</i>	<b>2,88</b>	<b>42,4</b>	<b>1,53</b>	<b>5,5</b>

Значимость полученных различий при сравнении четырех групп испытуемых проверялась по критерию Крускала и Уоллиса [1, стр. 229] при  $p < 0,001$ .

Входящие во вторую группу из четырех человек специалисты готовили цифровые данные для указанного «Журнала» и были знакомы с текстом в значительно меньшей степени.

В третьей группе из четырех человек специалисты совсем не были знакомы с угадываемым текстом, но так же хорошо, как и другие, были знакомы со специальным языком этого текста.

Четвертую группу из десяти человек составляли лица со средним образованием, не знакомые ни со специальным языком, ни с текстом.

Такой состав испытуемых позволял, с одной стороны, сравнить субъективную энтропию и избыточность специального текста для носителей данного языка, в разной мере знакомых с текстом и незнакомых с ним вообще; с другой стороны, можно было сравнить эти параметры для неспециалистов и специалистов (носителей специального языка) в случае, когда текст для тех и других был неизвестен.

Результаты опытов по угадыванию специального текста приведены в таблице.

Как показали опыты, количество информации в тексте, объективно очень информативном, субъективно невелико для специалиста, хорошо знакомого со специальным языком, даже в том случае, если текст ему заранее неизвестен.

Таким образом, большая информативность специальных языков сочетается с большой субъективной избыточностью для их носителей, что связано с очень ограниченным запасом слов в таких знаках. Это обстоятельство следует учитывать при решении проблемы общения человека с машиной.

#### ЛИТЕРАТУРА

1. Ж. Мот. Статистические предвидения и решения на предприятии. Изд-во «Прогресс», 1966.
2. П. Б. Невельский. Субъективная энтропия текста как ненадежность угадывания. Сб. «Проблемы языкознания. Докл. и сообщения советских ученых на X Международном конгрессе лингвистов, Бухарест, 1967». Изд-во «Наука», 1967.
3. Д. Ю. Панов. О проблеме общения человека с машиной. Сб. «Психологические проблемы системы человек — машина». (XVIII Международный психологический конгресс, Симпозиум 27). М., 1966.
4. Д. Ю. Панов. О взаимодействии человека и машины. «Вопросы философии», 1967, № 1.
5. Р. Г. Пиотровский. Информационные измерения языка. Изд-во «Наука», 1968.
6. К. Шеннон. Математическая теория связи. В кн. К. Шеннон. Работы по теории информации и кибернетике. Изд-во иностр. лит., 1963.
7. К. Шеннон. Предсказание и энтропия печатного английского текста. В кн. Работы по теории информации и кибернетике». Изд-во иностр. лит., 1963.

#### ВЫДЕЛЕНИЕ ПРОСТРАНСТВЕННЫХ ПРИЗНАКОВ ИЗОБРАЖЕНИЙ И КОДИРОВАНИЕ ЗНАКОВОЙ ИНФОРМАЦИИ

*С. Ф. Марченко, В. Г. Червов*

Для решения задачи распознавания необходимо выделять признаки, адекватные ее условию. При выделении этих признаков биологическими системами используются некоторые общие принципы, пригодные для решения любых задач. Физиологические и модельные исследования нервной системы и ее основной структурной единицы — нейрона показывают, что в процессе переработки поступающей информации из сообщений выделяются передаются по нервным волокнам только изменения наблюдаемого параметра. В основе такой обработки информации лежит процесс адаптации, присущий возбудимым образованиям на всех уровнях организации нервной системы [5]. Этот процесс обеспечивает поступление в высшие отде-