

УДК 811.111:81.367:81.23:81.32

МЕТОДЫ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ ИНФОРМАТИВНОСТИ ПАРАДИГМАТИЧЕСКОГО РЯДА СТРУКТУРНЫХ ФОРМ ЭМОЦИОНАЛЬНОЙ РЕЧИ

© 2003 Т.Н. Синеокова

Нижегородский государственный лингвистический университет
им. Н.А. Добролюбова

Статья посвящена описанию методов количественной оценки информативности (прогностического потенциала) парадигматического ряда форм эмоциональной речи. Разработанный ряд структурных модификаций адаптируется для целей прогноза эмоциональных состояний говорящего, оказывающих непосредственное воздействие на речемыслительные процессы - состояния диссоциации, состояния эвстресса и пограничного поискового состояния. Приводятся простейшие алгоритмы прогнозирования. Информативность классификации оценивается на основе уменьшения энтропии выбора по сравнению с априорными данными, содержащимися в общей эталонной выборке.

Любое исследование предполагает создание классификации как инструмента для описания наиболее существенных характеристик изучаемого объекта. Целью исследования синтаксических особенностей эмоциональной речи является создание такой классификации, которая позволила бы выявить корреляцию между психологическими состояниями говорящего и реализуемыми под их воздействием речевыми формами. Наряду с основной задачей разработки такой классификации встает задача тестовой проверки ее информативности. В качестве такой проверки, естественно, должно быть выбрано прогнозирование (идентификация) типа эмоционального состояния по структурным признакам высказывания. Смысл прогноза состоит в увеличении по сравнению с априорной вероятностью вероятности правильной идентификации типа состояния после соотнесения эмоционального высказывания с классификационными группами, то есть в снижении неопределенности выбора одного из возможных типов психологических состояний. Неопределенность выбора измеряется в математике параметром «энтропия выбора» (Больцмановская энтропия) [Стратонович, 1975]. Уменьшение энтропии определяет информацию (в дан-

ном контексте это числовой параметр), получаемую при соотнесении высказываний с классификационными группами, а процентное измерение энтропии можно считать числовой характеристикой информативности тестируемой классификации.

Предметом рассмотрения в настоящей статье являются опыт разработки классификации структурных форм эмоциональной речи и описание количественных методов оценки ее информативности на основе алгоритмов прогнозирования.

Как известно, в результате влияния эмоций на речемыслительные процессы возникают специфические формы, представляющие собой деформации структур, свойственных неэмоциональной речи. Такие формы можно представить в виде продуктивного парадигматического ряда вариантов некоего инварианта, характерного для нейтральной речи.

В качестве точки отсчета для построения парадигмы может служить типичное для нейтральной речи «ядерное предложение», характеризующееся четким набором дистинктивных признаков. Под воздействием психологического состояния говорящего признаки ядерного предложения деформируются, в результате чего возникает ряд модификаций (вариантов) с

новыми признаками, собственно и представляющий собой синтаксическую парадигму эмоциональной речи.

Разработка синтаксической парадигмы аффективной речи включает несколько этапов с последовательным возрастанием степени дифференциации признаков.

Первый ряд основан на дихотомическом подразделении всей специфики эмоциональной речи на экспликативные и имплекативные формы, второй представляет собой классификационный ряд наиболее общих структурных модификаций эмоциональных форм в рамках выделенных экспликативной и имплекативной реализаций, а третий включает в себя классификационный ряд более частных, «первичных» признаков внутри каждой из модификаций.

На первом этапе - этапе построения верхнего иерархического уровня парадигмы эмоциональной речи - представляется целесообразным выделить и классифицировать те наиболее общие признаки, которые, во-первых, отличают нейтральную языковую область от ненейтральной и, во-вторых, позволяют выявить наиболее характерные связи между ядерным инвариантом и его эмоциональными вариантами. Для решения этих вопросов представляется полезным использовать предложенное Ю.М. Скребневым дихотомическое подразделение всей коллоквиальной специфики на экспликационную и импликационную [Скребнев, 1961].

На следующем этапе разработки классификации - этапе построения второго иерархического уровня парадигмы эмоциональной речи - количественное измерение и последующая градация экспликативных и имплекативных форм осуществляется путем их сопоставления с признаками некоторой нейтральной единицы как точки отсчета. В качестве такой единицы может служить выделенное М.Я. Блохом «ядерное предложение», характеризующееся: 1) ограничением состава позиционным минимумом, отражающим лишь обязательные синтаксические позиции; 2) ориентированностью на валентностные

подклассы глагола, поскольку личный глагол является организующим центром предложения; 3) строго фиксированным порядком знаменательных позиций от подлежащего через личный глагол к предикативу, дополнению и обстоятельству; 4) максимальной коммуникативной независимостью структуры; 5) стилистически нейтральным характером структуры, соответствующим чистому типу «интеллективной», «рассудочной речи» [Блох, 1976 : 100]. Последний признак подчеркивает принадлежность перечисленных выше четырех дифференциальных признаков ядерного предложения к нейтральной области языка.

Изучение вариантов деформаций этих признаков позволяет выявить субпарадигматический ряд основных структурных форм эмоциональной речи, включающий такие модификации, как «повтор», «материально избыточные элементы», «инверсия», «транспозиция», «разрывы потенциального синтаксического целого», «расшифровка», «эллипсис» и «изолированные элементы» [см. об этом более подробно: Синеокова, 2001]. Таким образом, второй иерархический уровень субпарадигмы эмоциональной речи представляет собой классификационный ряд, состоящий из восьми модификаций, возникающих при деформациях признаков ядерного предложения. Выделенные модификации различаются формами проявления в них экспликационных и импликационных свойств. В одних случаях экспликационная и импликационная тенденции затрагивают конкретные материальные знаки (слова, словосочетания, потенциальные части сложных предложений), в других - структурно-семантические характеристики условий их объединения (порядок следования, слитность или разобщенность, функциональное значение конструкции и т.д.). Важной чертой ряда модификаций является то, что он содержит в себе возможность количественной градации признаков рассматриваемых структур и позволяет дать не только собственно лингвистическое, но и психологическое обоснование возник-

кающих построений. Это представляется особенно важным при изучении столь специфического объекта исследования, как синтаксис речи, порождаемой в состоянии эмоционального напряжения.

В то же время основным требованием к классификации является наличие у нее прогностического потенциала относительно существенного аспекта исследования. Как отмечает Н. Бейли, система классификации имеет объективную ценность лишь в том случае, если конкретное описание, которое она может обеспечить на основании имеющихся данных, позволяет осуществлять прогнозирование [Бейли, 1970]. Именно наличие такого потенциала определяет ценность и даже право на существование разработанной классификации. Предполагается, что классификация, обладающая хорошим прогностическим потенциалом относительно какого-либо важного аспекта, правильно отражает существенные особенности исследуемого материала и потому, скорее всего, окажется полезной и при прогнозировании других аспектов.

Полученный субпарадигматический ряд второго уровня, строго говоря, не является классификационным рядом, характеризующим исключительно эмоциональную речь, и не позволяет проводить прогнозирование. Входящие в него модификации могут возникать также вследствие намеренной или ненамеренной деформации ядерного предложения при логической эмфатизации высказывания или при разного рода затруднениях, не связанных с эмоциональным состоянием говорящего. Формализовать несомненно существующие различия синтаксиса деформаций различного происхождения пока не удастся. Для этого, по меньшей мере, необходимо исследовать синтаксис каждой из форм речи отдельно; таким образом, казалось бы, создается порочный логический круг. Однако интуитивно и тем более на основании анализа ситуации различить эти формы речи легко, и можно уверенно выделить речь эмоциональную. Поэтому важную задачу формализации признаков, позволяющих отграничить построения эмо-

циональной речи от других сходных построений, на данном этапе исследования можно не решать.

Кроме того, в пределах модификационных групп удобно выделить третий иерархический уровень общей парадигмы - субсубпарадигму первичных классификационных признаков, которые, в отличие от субпарадигмы модификационных групп, связаны не с общим эмоциональным состоянием говорящего, а с его конкретными проявлениями.

В то же время предварительная формальная проверка целесообразности использования субпарадигмы модификаций для ее дальнейшей разработки возможна. В качестве критерия промежуточного отбора может быть принято представление о «красоте», или «стройности», классификации с точки зрения исследователя. Исходными предпосылками для такой постановки вопроса являются следующие представления.

1. Существуют модели априорной информации о вероятностных закономерностях речи, которой владеет каждый взрослый индивид – носитель языка. Как замечает Р.М. Фрумкина, «вероятностная структура речевого опыта в основных своих чертах является общей для данной социально-культурной общности носителей языка и весьма стабильной во времени» [Фрумкина, 1996. 4 с.]. Порождая и воспринимая тексты, человек классифицирует несущие информацию элементы так, как это ему удобно.

2. Предпочтения исследователя при классифицировании элементов в основном совпадают с предпочтениями «среднего человека» и в этом смысле достаточно объективны.

3. Исходя из вышеназванных соображений, должна существовать классификация, которая с общечеловеческой точки зрения (и, следовательно, с точки зрения исследователя) является красивой истройной, т.е. удобной в работе. Ничего необычного в этих представлениях нет. Хорошую классификацию следует искать среди стройных классификаций, поскольку

ку классификация, отражающая глубинные характеристики объекта исследования, должна быть «красивой» (в технических ВУЗах, например, учат при конструировании машин «рисовать, как красиво», а потом вести проверочный расчет).

Были приняты следующие частные критерии качества («стройности») классификации: полнота охвата материала, ортогональность (непересекаемость) классификационных групп, число групп в каждой иерархии и равномерность распределения материала по группам. Для каждого из них был выбран числовой критерий, отражающий предпочтения исследователя.

Получившаяся шкала только упорядочивает оценки исследователя, но не измеряет их. Эти оценки аналогичны шкале твердости Мооса, балльной оценке знаний на экзамене или качества выступления фигуриста и т.п. Например, утверждение, что минерал с баллом 10 (например, алмаз) настолько же тверже минерала с баллом 8 (например, топаз), насколько минерал с баллом 5 (например, апатит) тверже минерала с баллом 3 (например, кальцит), совершенно бессмысленно, тогда как утверждение, что гиря с меткой 9 кг настолько же тяжелее гири с меткой 8, насколько гиря с меткой 5 тяжелее гири с меткой 4, имеет смысл. Вес - измеряемая величина, твердость (по Моосу) - нет.

Трудности возникают в основном при формировании интегральной оценки стройности. Ситуация аналогична определению общего балла по результатам вступительных экзаменов. При приеме, например, в Нижегородский лингвистический университет ценность знаний по английскому языку и по русскому языку не одинакова. Поэтому идеальному результату по английскому языку приписывается балл 20, а по русскому языку - 10. Коэффициенты пересчета (функции взвешивания частных оценок) должны быть подобраны и согласованы. Например, всегда можно достичь полной ортогональности классификации за счет значительного снижения полноты охвата или значительного увеличения числа групп. Согласова-

ние коэффициентов должно препятствовать получению кажущегося выигрыша от подобных бессмысленных операций. Методом проб на основе нескольких искусственно созданных, но интуитивно хорошо оцениваемых ситуаций подобный приблизительный подбор может быть сделан достаточно легко.

Существенно, что некоторый произвол, имеющий место в процессе таких оценок, мало критичен при выборе варианта-кандидата. «Красивых» вариантов мало, и они сильно отличаются от остальных. Например, типична ситуация, когда есть группа из двух фаворитов с интегральной оценкой 0.60 и 0.55 (абсолютное значение не существенно, существенны различия) и группа аутсайдеров с оценками 0.20, 0.17, 0.10. Возможные разумные изменения функций взвешивания могут поменять местами претендентов в группе фаворитов (и это ровно ничего не значит - все равно проверке подлежат оба претендента), но перевести вариант из группы фаворитов в группу аутсайдеров или наоборот - вряд ли.

Использование критериев «стройности» на промежуточном этапе разработки парадигмы эмоциональной речи - субпарадигмы эмоциональных модификаций - оказалось безусловно полезным. Интегральная оценка стройности субпарадигмы составила 0.53. Оценки стройности выделенного ряда модификаций составили: полнота охвата материала - 1, ортогональность - 0.66, число групп - 1, равномерность распределения материала по группам ряда - 0.80.

Возникает вопрос о возможности улучшения общей оценки классификационного ряда за счет повышения степени его ортогональности. Это может быть достигнуто только за счет многократного увеличения числа модификационных групп, что привело бы к усложнению классификации, снижению оценки по количеству групп и, соответственно, ухудшению интегральной оценки. По-видимому, достигнувшую степень ортогональности (0.66) можно считать приемлемой, поскольку наложение

признаков модификаций, лежащих в основе данной классификации, является характерной чертой эмоциональной речи и той данностью, которую надо не отвергать, «улучшая» классификацию, а принять, изучая характерные комбинации.

Следует отметить, что число групп (8) в предложенном субпарадигматическом ряде практически оптимально. Этот параметр был получен естественным образом и не потребовал каких-либо компромиссов в отношении других параметров стройности. Можно предположить, что такое положение не случайно; оно свидетельствует о том, что классификация, отражающая существующие особенности психики человека [см. о «магическом» числе 7 ± 2 : Miller, 1956], вероятно, отражает и другие закономерности, в том числе закономерности построения эмоциональной речи.

Таким образом, есть достаточно оснований рассматривать полученный субпарадигматический ряд как основу для дальнейшей разработки классификации по линии внутренней дифференциации каждой из модификационных групп. Полученные субсубпарадигмы представляют собой третий иерархический уровень общей парадигмы эмоциональной речи. Классифицирование первичных структурно-семантических признаков проводится в рамках экспликативных (повтор, материально избыточные элементы, разрывы потенциального синтаксического целого, расшифровка, инверсия) и имплективных (транспозиция, эллипсис, изолированные элементы) модификаций.

Разработанные субсубряды модификаций, состоящие из сорока первичных признаков, тестируются на полноту (способность описать любое эмоциональное высказывание) и однозначность (возможность для разных экспертов одинаково распознавать в исследуемом высказывании выделенные первичные признаки); используются также описанные выше формализованные параметры, характеризующие их «стройность».

Поскольку необходимым условием использования предложенного классифи-

кационного ряда для прогнозирования является возможность однозначной идентификации выделенных групп на основе их первичных признаков без обращения к авторским ремаркам и контексту, с целью проверки такой возможности было проведено тестирование. Всего было протестировано 520 высказываний (каждый эксперт тестировал 104 высказывания), отобранные случайным образом с помощью специальной компьютерной программы. Перед экспертами была поставлена задача разнести их по группам в соответствии с инструкцией, содержащей описание первичных признаков. Примеры, предъявляемые экспертам, не содержали авторских ремарок.

Результаты тестирования оценивались следующим образом:

1) относительное число ошибочных решений, принятых всей группой экспертов;

2) относительное число ошибочных решений, принятых экспертом, показавшим лучший результат;

3) относительное число ошибочных решений, принятых по мажоритарному принципу (учитывалось решение, принятое тремя и более экспертами).

Первые два критерия использовались в качестве оценки однозначности классификации. При этом первый критерий позволил определить возможный процент допускаемых ошибок в целом и, следовательно, подтвердить или не подтвердить однозначность выделенных первичных признаков. Исходной предпосылкой выделения второго критерия является то, что отсутствие однозначности не может быть компенсировано индивидуальными способностями эксперта. Третий критерий использовался для оценки качества формулировок и степени подробности инструкции. Таким образом, оценка качества классификации основывается на результатах, полученных лучшим экспертом, а оценка качества инструкции основывается на результатах, полученных худшим экспертом; мажоритарная оценка позволяет не учитывать результаты «необучаемых» экспертов.

Результаты тестирования дали следующий результат. Среднее количество ошибок составляет 3.3%, однако при мажоритарной оценке результатов было допущено менее 1% ошибок (1 ошибка на 104 высказывания). Три эксперта допустили 2.9% ошибок (3 ошибки в 104 высказываниях) и 2 эксперта - 3.8% ошибок (4 ошибки в 104 высказываниях).

Оценки субрядов по предложенным критериям стройности, как интегральные, так и частные, также достаточно высоки.

Таким образом, есть все основания для достаточно высокой оценки качества разработанной структурной классификации. Основной задачей исследования на следующем этапе является проверка наличия у нее прогностического потенциала относительно типа психологического состояния. Для этого необходимо провести адаптацию классификации применительно к прогнозированию психологических состояний говорящего, описать простейшие алгоритмы определения психологических состояний по структурным особенностям речи и количественно оценить возможности такого прогнозирования.

Поскольку с самого первого этапа разработка парадигмы эмоционального синтаксиса базируется на тезисе об определяющем влиянии психологического состояния говорящего на специфические речевые характеристики, представляющие собой отклонения реализуемой формы от нейтральной ядерной конструкции, создаваемые парадигматические ряды первичных признаков определенным образом коррелируют с классификационным рядом выделенных психологических признаков.

В качестве основы для построения психологического классификационного ряда рассматривались не сами эмоции, а более общий фактор: возникающие под их воздействием психологические состояния - «измененные состояния сознания», - оказывающие непосредственное влияние на речемыслительные процессы [Спивак, 1998] Среди таких состояний сознания были выделены три типа эмоциональных состояний, непосредственно связанных с

особенностями протекания речемыслительных процессов: состояние диссоциации, оказывающее деформирующее влияние на механизмы речепорождения (в дальнейшем XI), состояние эвстресса, стимулирующее речемыслительную деятельность говорящего (в дальнейшем X3), и поисковое состояние, являющееся пограничной фазой перехода от состояния диссоциации к состоянию эвстресса (в дальнейшем X2). Немаловажно отметить, что выделенные психологические состояния играют существенную роль не только для процесса речеобразования, но и для жизнедеятельности человека в целом.

При определении типов психологического состояния возможно использование авторских ремарок [о критериях отбора материала для анализа по авторским ремаркам см.: Ekman, 1965; Полищук, 1988; Синеокова, 1991; Солощук, 1990]. В некоторых случаях может потребоваться текстологический анализ на уровне элементарной тематической единицы связной речи - диктемы или диктемной группы - диктемемы [Блох, 2000]. Определенные таким образом психологические состояния используются как ключи при идентификации того или иного психологического состояния. Тестирование показало, что разные эксперты приблизительно одинаково идентифицируют по этим данным принадлежность высказывания к тому или иному состоянию. Существенно, что некоторое количество ошибок при определении контрольных ключей (что неизбежно при таком способе их идентификации) не оказывает существенного влияния ни на процесс разработки классификации, ни на разработку алгоритмов прогнозирования, поскольку ошибки нивелируются при усреднении. Эти ошибки несколько снижают численную оценку информативности классификации, так как их нельзя отличить от ошибок самого алгоритма; однако, поскольку прогнозирование все равно дает заниженную оценку информативности собственно классификации, то это тоже не очень существенно.

Саму задачу прогнозирования можно сформулировать следующим образом. Множество исследуемых элементов (в рассматриваемом случае это генеральная совокупность всех возможных эмоциональных высказываний) разделяется на подмножества - классификационные группы. Принадлежность высказывания к той или иной группе определяется по наличию в нем некоторых признаков (так называемых «первичных классификационных признаков»), доступных для наблюдения. В совокупности эти группы образуют классификационный ряд (классификацию). В соответствии с тем, в какую группу классификационного ряда попал тестируемый элемент, прогнозируются его другие, непосредственно не наблюдаемые свойства. Эта последняя стадия может быть формализована, т.е. может быть применен алгоритм принятия решения («алгоритм прогнозирования»), посредством которого определяется, какое из возможных прогнозируемых свойств присуще конкретному тестируемому элементу.

Для успешного прогнозирования необходимо преобразовать первичные классификационные группы таким образом, чтобы прогнозируемое свойство определялось ими, по возможности, однозначно, т.е. адаптировать классификацию к данному виду прогноза. Чтобы различать упомянутые два вида классификационных групп и рядов, последние названы прогностически адаптированными классификационными группами и рядами.

Таким образом, в процессе прогнозирования можно выделить три процедуры: выделение первичных признаков, формирование классификационных рядов разного уровня посредством многоступенчатого комбинирования признаков и групп и разработку алгоритма прогнозирования. Первая процедура является творческим процессом, не допускающим алгоритмизации, вторая допускает частичную алгоритмизацию, третья, естественно, полностью формализована. Поскольку алгоритм, как правило, включает в себя комбинирование классификационных групп, то граница

между второй и третьей процедурой несколько условна; в частности, формирование адаптивных групп может быть в некоторых случаях отнесено к алгоритму. Более того, существуют универсальные алгоритмы, базирующиеся непосредственно на первичных признаках и не требующие предварительного формирования адаптированных групп. По существу такие алгоритмы берут эту функцию на себя.

Разработка адаптированных групп рассматривается как самостоятельная процедура (выделенная из собственно алгоритма), поскольку формальные критерии при выделении групп часто оказываются недостаточными и приходится проводить анализ материала на содержательном (лингвистическом) уровне. Кроме того, нецелесообразно включать в состав алгоритма функцию формирования таких адаптированных групп, которые могут использоваться разными алгоритмами, даже если эти группы допускают формальное образование.

Прогностически адаптированные группы формировались с учетом их статистической значимости [Бронштейн, Семендяев, 1980]; принимались во внимание только те из них, которые включали достаточно большое число высказываний из эталонной выборки. Формирование адаптированных групп и контроль их статистической значимости позволил не только применить простые алгоритмы прогнозирования, но и использовать в качестве тестовой выборки для проверки их информативности ту же самую эталонную выборку, которая была использована при разработке классификации. Это позволило наиболее экономично использовать материал, подвергшийся весьма трудоемкой обработке, не создавая порочного логического круга. (Тем не менее, проводилось контрольное определение информативности по небольшой независимой тестовой выборке; статистически значимых расхождений при этих двух способах тестирования обнаружено не было.) Оказалось также, что выделенные адаптированные группы хорошо «высвечивают» ряд закономерностей син-

таксиса эмоциональной речи, являясь своеобразными функционально синонимическими парадигматическими рядами. Отмеченные при этом закономерности были затем дополнительно исследованы и подтвердились.

По отношению к каждому из трех психологических состояний были выделены следующие адаптированные группы: «практически 0», «практически 100%», «обедненные», «обогащенные» и «средние». При этом создавались как частные прогностически адаптированные группы, разработанные на основе каждой из восьми модификаций, так и обобщенные прогностически адаптированные группы, включающие в себя весь корпус признаков.

При тестировании использовались и сравнивались между собой четыре алгоритма прогнозирования.

Первый алгоритм. Анализируются только обобщенные группы («практически 100%»). В тех случаях, когда пример принадлежит только к одной такой группе, принимается решение об идентификации его как психологическое состояние X, соответствующее этой группе. В редких (поскольку ряд почти ортогонален) случаях, когда пример принадлежит одновременно двум или более группам, конфликт разрешается следующим образом: группа, соответствующая состоянию X1, имеет приоритет над группами, соответствующими состояниям X2 и X3, а группа, соответствующая состоянию X2, имеет приоритет над группой, соответствующей состоянию X3. Принадлежность примера к обогащенным группам вообще не учитывается. Этот алгоритм способен идентифицировать около 40% всех примеров с очень небольшим процентом ошибок.

Второй алгоритм. Анализируются все шесть обобщенных групп («практически 100%» и «обогащенные»). Анализ проводится в два этапа. На первом этапе выполняется алгоритм 1. Если он дает результат (пример попал хотя бы в одну группу «практически 100%» и, следовательно, идентифицирован), анализ заканчивается.

Если нет - проводится второй этап: анализируются только обогащенные группы. Разрешение конфликта производится аналогичным образом, то есть группа, соответствующая состоянию X1, имеет приоритет над остальными и т.д. Таким образом, при этом алгоритме имеет место приоритет любой группы «практически 100%» над любой «обогащенной», а внутри однородных групп действует приоритет по идентифицируемому психологическому состояниям. Этот алгоритм способен идентифицировать около 80% всех примеров, но число ошибок при этом возрастает.

Третий алгоритм. В этом алгоритме используются как общие, так и частные прогностически адаптированные группы. Сначала, как и в алгоритмах 1 и 2, рассматриваются обобщенные группы. Анализируются все шесть групп. Анализ проводится в два этапа. На первом этапе выполняется алгоритм 1. В случаях, когда пример попадает в две обобщенные адаптированные группы одновременно, конфликт разрешается следующим образом: для каждой из конкурирующих обобщенных групп подсчитывается число стопроцентных частных групп, в которые входит этот пример. Приоритет отдается обобщенной группе, которой соответствует наибольшее количество частных адаптированных групп. Если это число одинаково, то приоритет определяется в зависимости от идентифицируемого эмоционального состояния. Если первый этап дает результат (высказывание попало хотя бы в одну группу «практически 100%»), анализ заканчивается. Если нет - проводится второй этап: таким же образом, как на первом этапе, анализируется только обогащенная группа. Разрешение конфликта производится аналогичным образом, то есть в два приема: сначала по числу частных обогащенных групп, а затем в соответствии с приоритетом эмоционального состояния. Этот алгоритм способен идентифицировать столько же примеров, сколько и второй, но распределение ошибок может быть несколько иным.

Четвертый алгоритм. Определяется принадлежность тестируемого высказывания к каждой из 15 обобщенных прогнозистически адаптированных групп. Высказыванию присваиваются следующие оценки: в случае принадлежности к группе «100%» - $K_5=10$, к «обогащенной» группе - $K_4 = 3$, к «средней» - $K_3 = 0.3$, к «обедненной» - $K_2 = 0.3$, к группе «практически 0» - $K_1=0.1$. В случае, если высказывание не принадлежит к тестируемой группе, в качестве соответствующей оценки принимается 1. Для каждого состояния X_i вычисляется интегральная оценка $K_{xi} = K_1 * K_2 * K_3 * K_4 * K_5$ (по группам ряда, адаптированного к состоянию X_i). Определяется максимальная интегральная оценка, и, если она превосходит 1, высказывание идентифицируется в соответствии с состоянием, к которому она относится (с приоритетом $X_1 - X_2 - X_3$ для разрешения конфликта). Если максимальная интегральная оценка меньше или равна 1, то высказывание не идентифицируется. Этот алгоритм способен идентифицировать несколько меньше высказываний, чем второй и третий, за счет отказа от идентификации части «сомнительных» примеров. Реализация алгоритма требует в 3-4 раза больше машинного времени, чем остальные, но дает значительно меньше ошибок.

Следует отметить, что все четыре рассмотренных алгоритма в известной мере учитывают не только предпочтения, отдаваемые исследователем обнаружению различных типов того или иного психологического состояния, но и различную оценку значимости ошибок разного рода (пропуска, ложной индикации). Последнее обеспечивается, в частности, вариацией границ «обогащенных» и «практически 100%» групп для разных состояний. Очевидно также, что ценою усложнения алгоритмов можно улучшить результат прогнозирования. Однако разработка оптимального алгоритма прогнозирования не входила в задачи работы, и для тестирования были отобраны простейшие типы.

Все примененные алгоритмы прогнозирования дали примерно одинаковую

оценку информативности комплекса «классификация – алгоритм»: между 0.44 и 0.47 при возможном максимуме 1. Характер примененного для оценки параметра (% изменения энтропии) таков, что этот результат можно считать значительным.

На самом деле определенная таким образом информативность не учитывает «ценности» информации. Очевидно, что правильная идентификация редкого состояния X_1 , составляющего 6% в эталонной выборке, ценнее, чем правильная идентификация состояния X_3 , составляющего 82% выборки. Если рассматривать состояние X_1 как «опасное» (а это так и есть), то очевидно, что цена ошибки типа пропуск («пропуск угрозы») больше, чем цена ошибки типа ложное обнаружение («ложная тревога») и т.д. Для полного описания результатов тестирования необходимо использовать 12 параметров, а интегральная оценка информативности дает лишь общее представление. Ситуация напоминает шутку о «средней температуре по госпиталю». Но недооценивать подобные средние параметры, видимо, тоже не стоит: средняя температура 18°C говорит о том, что это скорее морг, а 40°C - о тяжелой эпидемии. Действительно, результат, когда редкие состояния X_1 и X_2 идентифицируются в 82% и 70% случаев практически без ошибок, можно считать достаточно хорошим. Подчеркнем, что прогнозирование проводилось исключительно на основе структурных характеристик высказывания, без учета специфических особенностей эмоциональной речи на фонетическом и лексическом уровнях.

Полученные высокие оценки качества классификации и успешное прогнозирование позволили применить разработанную классификацию для выявления и описания некоторых статистических закономерностей реализации структур в эмоциональной речи. Данные закономерности затрагивают частотность, разнообразие, сложность, доминантность и сочетаемость реализуемых в ней построений.

Таким образом, проведенный анализ подтверждает, во-первых, возможность

разработки прогностически информативного парадигматического ряда специфических структур эмоциональной речи и, во-вторых, его пригодность для целей анализа общих свойств эмоциональной речи. Кроме того, представляется, что предлагаемая методология может быть успешно использована при решении таких вопросов, как вопросы разграничения эмоциональных форм и сходных с ними конструкций, возникающих вследствие намеренной модификации или ненамеренной деформации ядерного предложения при неаффективной эмфатизации высказывания или языковых затруднениях, не связанных с эмоциональным состоянием говорящего, вопросы определения прогностического потенциала выделяемых структурно-семантических признаков в отношении других внелингвистических характеристик говорящего (его социального статуса, пола, возраста и т.п.) и, наконец, вопросы специфики эмоциональных форм речи в разных языках.

СПИСОК ЛИТЕРАТУРЫ

1. Бейли Н. Математика в медицине и биологии. М., «Мир», 1970.
2. Блох М.Я. Проблемы парадигматического синтаксиса (На материале английского языка): Дис. ...докт. филол. наук. М., 1976.
3. Блох М.Я. Диктема в уровневой структуре языка // Вопросы языкознания. 2000. № 4.
4. Бронштейн И.Н., Семендяев К.А. Справочник по математике для инженеров и учащихся втузов. М.: Наука, 1980.
5. Полищук Н.В. Номинативный статус междометных фразеологических единиц современного английского языка и особенности их контекстного употребления: Дис. ... канд. филол. наук. М., 1988.
6. Синеокова Т.Н. Индикаторы аффекта в авторских ремарках (на материале

пьес современных английских и американских драматургов) // Теория и практика лингвистического описания разговорной речи: Межвуз. сб. научн. тр. - Нижний Новгород: НГПИИЯ им. Н.А. Добролюбова, 1991.

7. Синеокова Т.Н. Парадигматический подход к изучению синтаксиса эмоциональной речи // Владимир Даль и современная филология: Материалы межвузовской научной конференции. 22-23 ноября 2001 года. Нижний Новгород. Том I. - Нижний Новгород: Нижегородский государственный лингвистический университет им. Н.А. Добролюбова, 2001.
8. Синеокова Т.Н. Общие статистические характеристики синтаксиса эмоциональной речи // Язык. Речь. Речевая деятельность: Межвузовский сборник научных трудов. Выпуск пятый. - Н. Новгород: Нижегородский государственный лингвистический университет им. Н.А. Добролюбова, 2002.
9. Скребнев Ю.М. К вопросу об «ортологии» // Вопросы языкознания. 1961. № 1.
10. Солощук Л.В. Структурно-семантические и прагматические характеристики ремарок в драматургическом произведении: (На материале английского языка): Дис. ... канд. филол. наук. Киев, 1990.
11. Спивак Д.Л. Лингвистика измененных состояний сознания: Дис. ... докт. СПб., 1998.
12. Стратонович Р.Л. Теория информации. М.: Советское радио. 1975
13. Фрумкина Р.М. «Теории среднего уровня» в современной лингвистике // Вопросы языкознания. 1996. ¹ 2.
14. Ekman P. Differential Communication of Affect by Head and Body Cues // Journal of Pers. Soc. Psychol. - 1965. V.2.
15. Miller G. The magical number seven, plus or minus two // Psychol. review. 1956. V. 63.

QUANTITATIVE ASSESSMENT METHODS FOR THE INFORMATIVE CONTENT OF STRUCTURAL FORMS IN EMOTONAL SPEECH

© 2003 T.N. Sineokova

Nizhni Novgorod Linguistic University named after N.A. Dobrolubov

The paper discusses quantitative assessment methods for the informative content of structural forms in emotional speech. Modifications of emotional syntactic paradigm are adapted for the purposes of prediction of emotional states affecting speech-thought processes, such as dissociation, eustress and the borderline searching state. Prediction algorithms are proposed. Validity of the paradigmatic classification is evaluated on the basis of decreasing entropy of choice in comparison with the prior data in the reference sample.