

Рассмотрев переводческие трансформации в исследуемых текстах, мы пришли к выводу, что наиболее характерными для данных текстов являются конкретизация, адекватная замена и логическое развитие. Иногда встречаются также замена частей речи и генерализация.

ЛИТЕРАТУРА

1. *Шеремет, О.* Специфика формирования и использования параллельных корпусов текстов / О. Шеремет // Материалы ежегод. науч. конф. студентов и магистрантов ун-та, 27–28 апреля 2010г. : в 2 ч. / Минский гос. лингвист. ун-т; отв. ред. Н. П. Баранова. – Минск, 2010. – Ч. 2. – С. 81–82.
2. *Степанова, Д. В.* Возможные способы установления соответствий между единицами параллельных текстов / Д. В. Степанова // Материалы ежегод. научн. конф. преподавателей и аспирантов ун-та, 24–25 апр. 2007 г. : в 5 ч. / Минский гос. лингвист. ун-т; редкол.: Н. П. Баранова (отв. ред.) [и др.]. – Минск, 2007. – Ч. 4. – С. 134–138.
3. *Шеремет, О.* Параллельные тексты и особенности их исследования / О. Шеремет // Материалы ежегодной научной конференции студентов и магистрантов ун-та, 21–22 апр. 2009 г. : в 2 ч. / Минский гос. лингвист. ун-т; отв. ред. Н. П. Баранова. – Минск, 2009. – Ч. 2. – С. 46–47.
4. *Федоров, А. В.* Основы общей теории перевода (Лингвистические проблемы) : учеб. пособие / А. В. Федоров. – 5-е изд. – СПб. : Фил. фак-т СПбГУ. – М., 2002. – 416 с.
5. *Шеметов, В. Б.* Курс лекций по теории перевода для студентов второго и третьего курсов, а также до подготовки к гос. экзаменам / В. Б. Шеметов. – М., 1998. – 60 с.
6. *Комиссаров, В. Н.* Современное переводоведение : курс лекций / В. Н. Комиссаров. – М. : ЭТС, 1999. – 192 с.
7. *Комиссаров, В. Н.* Теория перевода: лингвист. аспекты / В. Н. Комиссаров. – М. : Высш. шк., 1990. – 250 с.

The article deals with the notion of "parallel text", as well as the translation process optimization with the help of parallel texts. The results and main conclusions of the research of lexical transformations in the automatic translation of English texts of 17 UN goals for sustainable development into Russian are presented.

А. В. Скрыбина

Минск, МГЛУ

АВТОМАТИЗАЦИЯ РАСПРЕДЕЛЕНИЯ АНГЛОЯЗЫЧНЫХ НОВОСТНЫХ ТЕКСТОВ ПО ТЕМАТИКАМ

В статье рассматриваются некоторые аспекты автоматической тематической классификации англоязычных новостных текстов.

Количество интернет-ресурсов, посвященных публицистическим текстам, растет с каждым днем. Именно поэтому довольно актуальной является разработка автоматической системы тематической классификации публицистических текстов. В данной работе нами была предпринята попытка создания такой системы для классификации отобранного лингвистического материала. Источниками послужили англоязычные интернет-ресурсы, посвященные публицистическим текстам, такие как www.bbc.com и www.theguardian.com. Всего для анализа нами было отобрано 80 англоязычных новостных текстов по теме «культура». Далее отобранные тексты по тематике «культура» были разделены на десять более узких тематик на основании тематической принадлежности публицистических текстов.

В отобранном нами материале тексты представлены довольно однородно. Наиболее многочисленная группа из монотематических текстов – «Film» (15 %). Также больше всего текстов по тематике «Music» (13 %), «Theater» (13 %).

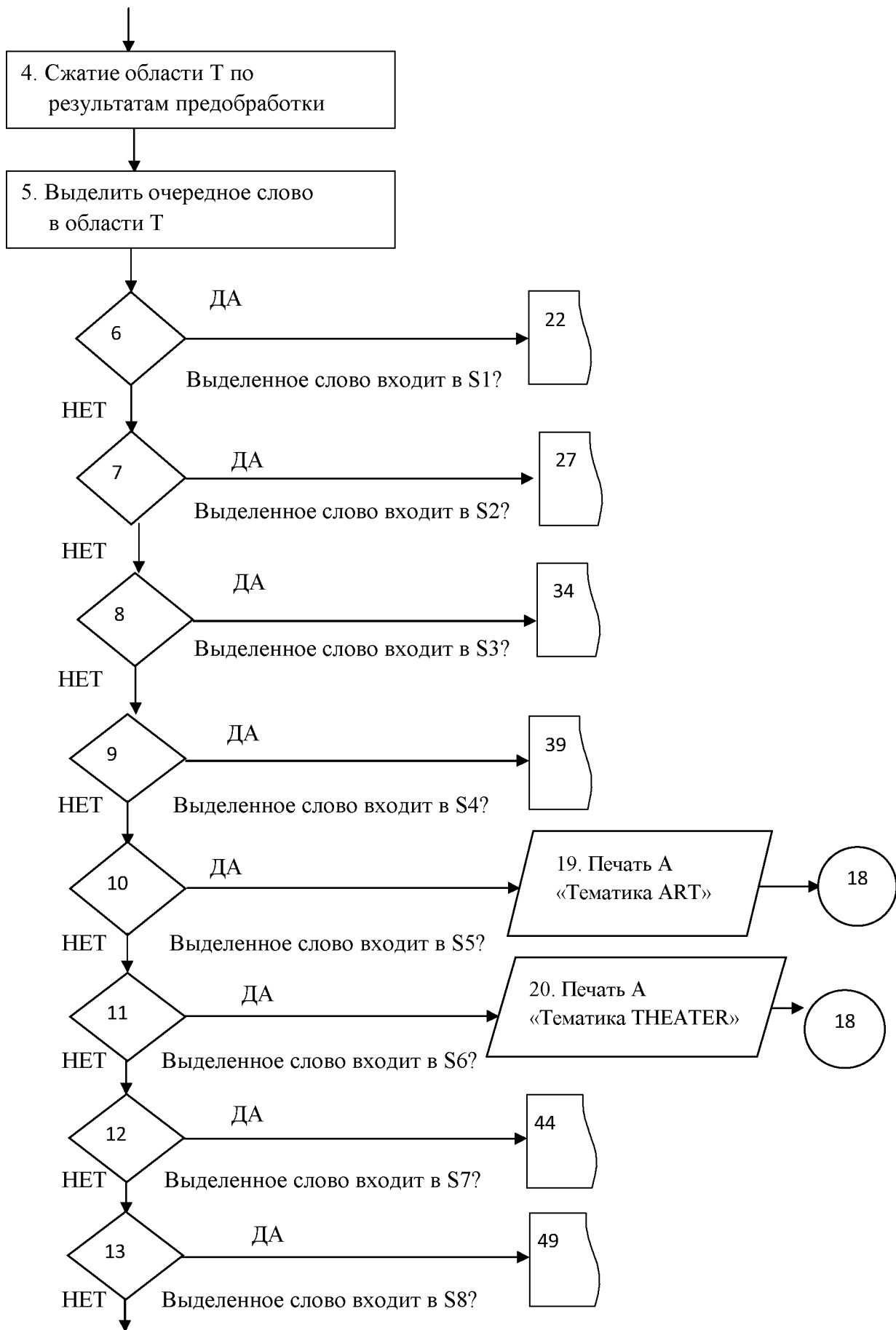
Политематические тексты составляют вторую по численности группу (15 %). В эту группу вошли такие тематики как: «TV & Dance» (10), «TV & Music» (16), (30), «Photography & Books» (32), «Photography & Film» (27), «Film & Fashion» (35). В результате мы получили 12 публицистических текстов, которые объединяют 2 тематики.

На основании разработанных поисковых образов по каждой тематике был составлен алгоритм автоматической идентификации тематики англоязычных новостных текстов (фрагмент алгоритма представлен на рис. 1).

Для наглядности и удобства работы при разработке алгоритма мы использовали условные обозначения для поисковых образов по каждой тематике:

- S1 – поисковый образ для тематики «BOOKS»;
- S2 – поисковый образ для тематики «FILM»;
- S3 – поисковый образ для тематики «DANCE»;
- S4 – поисковый образ для тематики «MUSIC»;
- S5 – поисковый образ для тематики «ART»;
- S6 – поисковый образ для тематики «THEATER»,





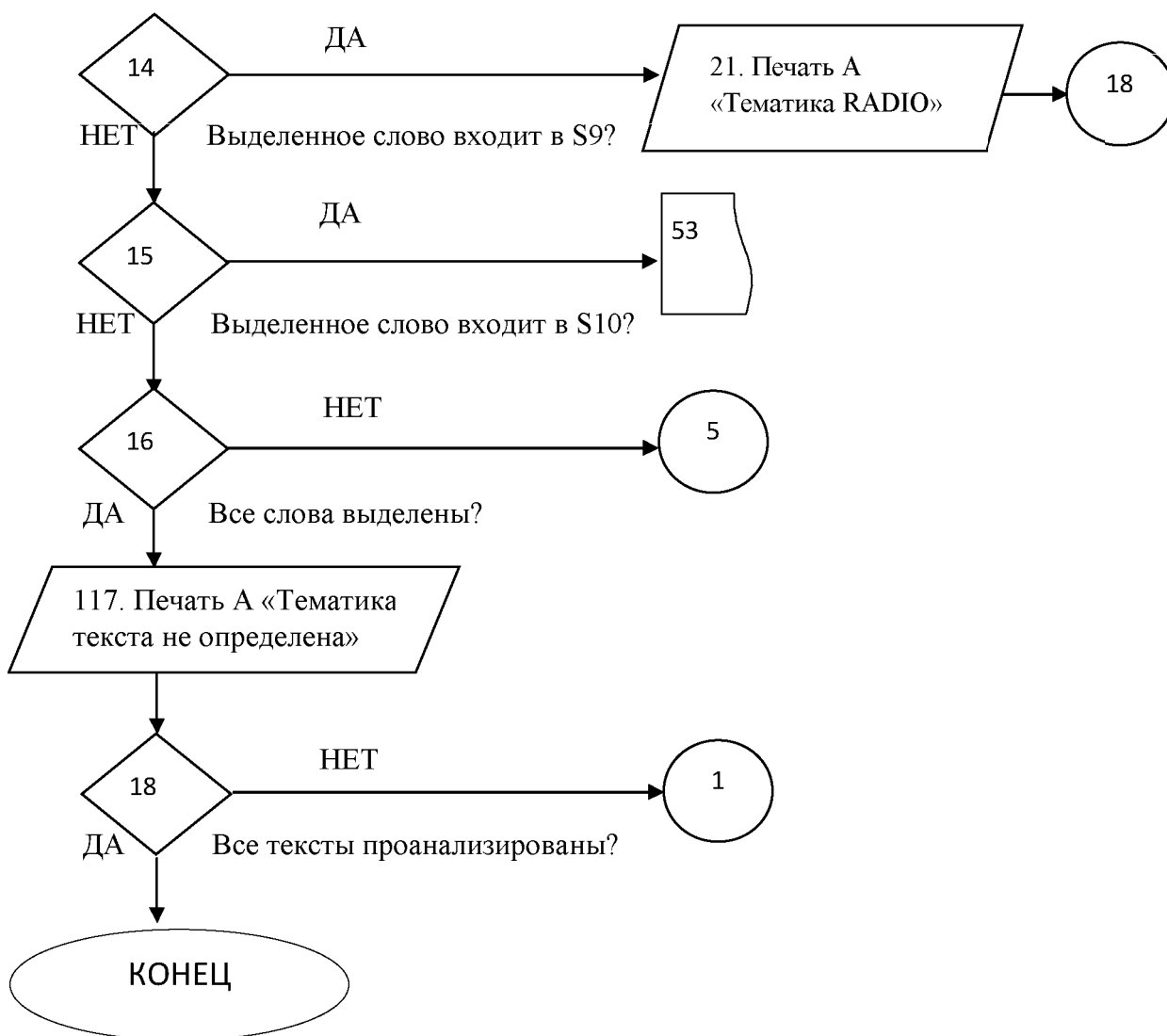


Рис. Алгоритм тематической классификации англоязычных публицистических текстов (фрагмент)

S7 – поисковый образ для тематики «TV»;

S8 – поисковый образ для тематики «FASHION»;

S9 – поисковый образ для тематики «RADIO»;

S10 – поисковый образ для тематики «PHOTOGRAPHY».

Процесс индексирования и создания поискового образа документа является актуальным в области компьютерной лингвистики в настоящее время. Эти системы особенно необходимы для обработки публицистических текстов, так как на данный момент существует большое количество таких текстов. Обработать новостные тексты вручную и выделять их основное содержание – процесс для человека довольно длительный и трудоемкий. Поэтому выделение основного содержания текстов и создание современных систем автоматического отнесения новостных текстов к определенным тематикам является востребованным в наши дни.

The article considers some aspects of automated thematic classification of English news texts.