

Імя Піліп сустракаецца ў ФА як піліп з канапель «зусім нечакана, знянацку». Часам узнікненне фразеалагізма звязваюць (услед за П. Шпілеўскім) з наступнай легендай: аднойчы на сейміку магната Радзівіла нечакана і недарэчна выказаўся шляхціц Піліп з Канапель (так называўся засценак на Чэрвеньшчыне), за што быў востра абсмяяны. Аднак існуе іншае тлумачэнне: у польскай мове слова *filip* — народная назва зайца. Магчыма, выраж як піліп з канапель склаўся ў асяроддзі паляўнічых, якія не раз бачылі падчас палявання, як, забегшы ў каноплі ў перыяд іх цвіцення, заяц-піліп не вытрымліваў іх ап'янняльнага паху і стрымгалоў высококаў адтуль [3; 429].

Імя Эзоп выкарыстоўваецца ў ФА эзопаўская (эзопава) мова «іншаскальнае, алегарычнае выкаванне думак», якая з'яўляецца паўкалькай з нямецкай ці французскай мовы, яе ўзнікненне звязана з літаратурнай дзеянасцю старожытнагрэческага байкапісца Эзопа (VI ст. да н.э.), які не мог свабодна і адкрыта гаварыць пра многае, бо гэта было небяспечна (ён быў рабом), таму выказаў свае думкі ў алегарычнай, баечнай форме [3; 412].

Як відаць з прыкладаў, у аснову асобных ФА пакладзены рэальныя факты і падзеі.

Такім чынам, ФА могуць утрымліваць уласныя асабовыя імёны як рэальных асоб, шырока вядомых грамадству (Мамай, Эзоп, Варфаламей), так і найменні выдуманых фантазій байкапісца персанажаў (Пётр, Дзям'ян). Антрапонімы становяцца неад'емнай часткай такіх устойлівых адзінак, неабходным і дарэчным сродкам характарыстыкі і выражэння адносін.

Літаратура

1. Лепешаў, І. Я. Фразеалагічны слоўнік беларускай мовы / І. Я. Лепешаў. — Мінск : Беларус. Энцыкл., 1993. — Т. 1: А–Л. — 590 с.
2. Лепешаў, І. Я. Фразеалагічны слоўнік беларускай мовы / І. Я. Лепешаў. — Мінск : Беларус. Энцыкл., 1993. — Т. 2: М–Я. — 607 с.
3. Лепешаў, І. Я. Этымалагічны слоўнік фразеалагізмаў / І. Я. Лепешаў. — Мінск : Беларус. Энцыкл., 2004. — 448 с.

Понятие об именованной сущности в письменном тексте

Мурашко В. Н., студ. V к. МГЛУ,
науч. рук. Зубова И. И., доц.

Английский термин *Named Entity (NE)* — именованная сущность — был впервые использован в 1995 г. в рамках конференции *MUC (Message Understanding Conferences)* в связи с постановкой задачи автоматического распознавания именованных сущностей [1]. Под именованной сущностью понимают «что-либо реально существующее или вымышленное, на что

можно указать или к чему можно обратиться при помощи имени собственного» [2; 214]. Таким образом, именованная сущность обязательно имеет референт, то есть объект внеязыковой действительности, подразумеваемый автором конкретного речевого отрезка. Референт может принадлежать реальному миру, например, фамилия, имя и отчество конкретного человека или название конкретной организации, либо вымышленному миру, например, это может быть персонаж художественного произведения.

По мнению ряда ученых в письменном тексте именованные сущности представлены объектами и их атрибутами [3; 294–295]. Под объектами понимаются либо персоны, имеющие такие атрибуты, как фамилия, имя, отчество, должность, звание и т. д., либо организации и топонимы, характеризующиеся такими атрибутами, как представленное в полном или сокращенном виде название предприятия, его адрес, направление деятельности и т. п. Выделяют три типа атрибутов. Во-первых, атрибуты, характерные непосредственно для текста документа, например, заголовок, дата создания, исполнитель документа; во-вторых, атрибуты, характерные исключительно для лиц, например, год рождения, национальность, особенности внешности человека и, в-третьих, атрибуты, свойственные двум отмеченным выше типам признаков, например, адрес, номер и марка машины.

В соответствии с приведенным выше определением в задачу распознавания именованной сущности входит не только нахождение ее в тексте, но и однозначное указание на подразумеваемый объект или лицо, а также приписывание ей определенной категории. Чаще всего, используется простая классификация, включающая в себя следующие три категории: ЧЕЛОВЕК, МЕСТО и ОРГАНИЗАЦИЯ [4; 1538]. Хотя ни даты, ни числа не соответствуют определению понятия именованной сущности, они часто распознаются, как таковые и считаются, соответственно, темпоральными и нумерическими выражениями [4; 1539]. Следует отметить, что, в отличие от трех других категорий, распознавание чисел и времени представляется значительно более легкой задачей, поскольку существует ограниченный набор способов их выражения в тексте.

Рассмотрим взятый из Википедии пример, содержащий разные именованные сущности и соответствующие им категории: *L'université de Strasbourg* est une université française située à [Strasbourg] en [Alsace]. Son origine remonte à la création du [Gymnase Jean-Sturm] en [1538]. Le Gymnase fut promu au rang d'Académie en [1566], sous l'empereur [Maximilien II], puis devint université en [1621], sous l'empereur [Ferdinand II] [5]. Соотнесем выделенные жирным шрифтом именованные сущности с возможными категориями. Так, именованную сущность *université de Strasbourg* можно отнести к категориям *организация, образовательное учреждение, вуз*; именованную

сущность *Gymnase Jean-Sturm*, соответственно, к категориям *организация, образовательное учреждение, гимназия*; *Strasbourg* — к категориям *место, город*; *Alsace* — к категориям *место, регион*; *1538, 1566, 1621* — к категориям *время, дата*; *Maximilien II и Ferdinand II* — к категориям *человек, исторический деятель, политик, правитель*. Как видно из примера, помимо очевидной категории «человек», персонам *Maximilien II* и *Ferdinand II* можно приписать и другие категории, соответствующие этим многогранным личностям. В данном случае речь идет не о более подробной детализации, как в случае с *université de Strasbourg* (от организации к учебному заведению и дальше к вузу), а о категориях одного порядка. Исходя из вышеизложенного, нередко возникает вопрос о том, нужно ли проводить границу между двумя сущностями или их можно считать единым объектом. Возвращаясь к приведенному выше примеру, рассмотрим единицы *Université de Strasbourg* и *Gymnase Jean-Sturm*. Из текста следует, что один является преемником другого и что именно это дает возможность говорить о том, что дата основания *Université de Strasbourg* — 1538 г. С другой стороны, признавая, что *Université de Strasbourg* и *Gymnase Jean-Sturm* — одно и то же заведение, можно лишить текст смысла, потому что утверждение *Université de Strasbourg является преемником Université de Strasbourg* не несет в себе никакой информации. В приведенном контексте это два разных заведения (две разные именованные сущности), однако в каких-то других контекстах можно считать, что речь идет об одном и том же заведении. Таким образом, выбор соответствующей категории не обязательно следует из именованной сущности. Очень часто категория определяется контекстом, в котором сущность была упомянута.

Литература

1. Кузнецов, И. Методики выявления объектов и связей, заданных в неявном виде [Электронный ресурс]. — Дата доступа: 18.11.2017. — Режим доступа: <http://www.dialog-21.ru/digests/dialog2012/materials.pdf>.
2. Прикладная и компьютерная лингвистика / под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо. — М. : ЛЕНАНД, 2016. — 320 с.
3. Котельников, Д. С. Итерационное извлечение шаблонов описания событий по новостным кластерам / Д. С. Котельников, Н. В. Лукашевич // Труды 14-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL–2012, Переславль-Залесский, 15–18 октября 2012 г. — Переславль-Залесский, 2012. — С. 292–298.
4. Fader, A. Identifying Relations for Open Information Extraction / A. Fader, S. Doderland, O. Etzioni // Proc. of EMNLP-11. — 2011. — Р. 1535–1545.
5. Википедия [Электронный ресурс]. — Дата доступа: 24.11.2017. — Режим доступа: <http://www.wikipedia.org>.