

## Chapter 2

# Semantic and syntactic patterns of multiword names: A cross-language study

Svetla Koeva

Institute for Bulgarian Language, Bulgarian Academy of Sciences

Cvetana Krstev

Faculty of Philology, University of Belgrade

Duško Vitas

Faculty of Mathematics, University of Belgrade

Tita Kyriacopoulou

Université Paris-Est Marne-la-Vallée

Claude Martineau

LIGM, Université Paris-Est Marne-la-Vallée

Tsvetana Dimitrova

Institute for Bulgarian Language, Bulgarian Academy of Sciences

Named entities (NEs) constitute a great challenge for computational linguistics and one of the major research topics during the last decade. They can be divided in categories describing people, location, time, organization and others. In this paper we will restrict our discussion to proper names that belong to three main classes: personal, location and organization names, and that can be either single-word nouns or multiword expressions. First, we are going to define common (language-independent) semantic patterns for proper names and then we will present the cor-



responding syntactic patterns in English, Bulgarian, French, Greek, and Serbian. We will compare these patterns regarding grammatical categories of dependent constituents, definiteness, distribution of clitics, word order and various alternations. Our ultimate goal is to build a universal framework for Named Entity Recognition (NER).

## 1 Introduction

Proper names are usually defined as belonging to the following main classes: personal names, location names, and organization names, called also named entities (NEs). They can be single-word nouns or particular types of multiword expression (MWE).

The aim of this paper is to offer a common template for description and classification of proper names in different languages. Our objectives are: i) to formulate semantic patterns for personal, location and organization names that capture the general semantics and should be, to a great extent, language-neutral; ii) to describe language-specific syntactic patterns corresponding to a common semantic pattern. The syntactic patterns provide information about the grammatical class of the head and constituents; dependencies among the constituents; word order and contiguity; cliticisation (if applicable – for possessive pronoun and interrogative clitics).

This study is based on evidence gathered from five languages – English, Bulgarian, French, Greek, and Serbian – belonging to four different language groups (Germanic, Hellenic, Romance, and Slavic). The utility of language-neutral semantic patterns lies in the fact that they can be applied to new languages, thus paving the way towards more universal solutions for (rule-based) named entity recognition (NER). The set of language-specific syntactic patterns displays correspondences between morphological and syntactic language-specific characteristics, and they may serve as transformation rules in rule-based machine translation, cross-lingual information extraction and summarization.

## 2 Names: a general overview

In English (Huddleston 1988: 96), two different terms are often used: *proper noun* – referring to the part-of-speech of the word and comprising only single-word proper names, e.g., *John, London, Adidas*, and *proper name* – referring to the function of these words as referential elements and comprising single- and multiword proper names, as in *John, John Smith Junior, London, the United States of America*,

*Nike, Microsoft Corporation*. Following this distinction, proper names can be further specified as: proper nouns (*Anna, Asia, Google*), multiword expressions (*Jean-Pierre Deckles, New York, the United Nations*), and noun phrases (*Professor Deckles, New York City, the United Nations Organization*). Proper names – expressed either by proper nouns or by MWEs – show common semantic and syntactic behavior and we describe them in a uniform way.

Proper names do not “[d]escribe or specify characteristics of objects” but are “logically connected with characteristics of the object to which they refer” (Searle 1958: 173). For example, *Saint Petersburg* may refer to the second largest city in Russia (*They convened in Saint Petersburg*), a city in Florida, a city in Pennsylvania, the fictional hometown of Tom Sawyer and Huckleberry Finn, *St. Petersburg, Missouri*, but also to a college in Florida – *Saint Petersburg College*. A particular common noun (i.e., *city, street, president, actor*) specifies the object whose instances may be represented by a set of proper names; and such an object is always presupposed for a given proper name even if the common noun is not mentioned explicitly in the text. Furthermore, multiword names may comprise a category word (*square* in *Trafalgar Square*; *ocean* in *the Indian Ocean*), and in these cases, the category of the particular name is always explicitly shown (Carroll 1985: 144).

The relation between a proper name and its category object is reflected in WordNet where the relation between a concept and its instances is defined as an instant hypernym (instant hyponym) relation (Rodríguez et al. 1998). The instances (proper nouns) inherit characteristics from the concepts of the hierarchy to which they belong. For example, the name *Saint Petersburg* is an instance of a city, and the concept {city, metropolis, urban centre} links to the more general concept {region}, which, in its turn, links to an even more general concept {location}.

Different terms are used for common nouns that categorise the referents of proper names as members of different classes: descriptors, designators, category words (Carroll 1985), external evidence (McDonald 1996), triggers (Magnini et al. 2002), trigger words. To avoid confusion with the theory of reference, we will use the term *trigger*.

Triggers depend semantically on the referent of the personal name, and different names select different classes of triggers. In turn, triggers determine the characteristics of the object to which the name refers. For example, if we know that the word *Washington* is a family name, it can select the word *president* or the word *actor*. Further, the word *president* and the word *actor* are similar in the way they designate the concept for a person, and this determines the fact that

both nouns can co-occur with adjectives denoting height, age, etc. The meaning of both words also implies that they may be specified by employing expressions for affiliation as complements (*the President of the USA, the actor at the Muppet Theatre*). However, not all words that are compatible with the first noun are compatible with the second (*stage actor* vs. \**stage president*). Therefore, the notion of triggers is central for the classification of the semantic patterns of proper names and accordingly – for the description of the respective syntactic patterns.

### 3 Grammatical features of names in Bulgarian, English, French, Greek and Serbian: a brief overview

Personal names are singular and inherently definite (the same applies to location and organization names that cannot express definiteness). Some Bulgarian, English, French, Greek, and Serbian location and organization names are in singularia tantum or pluralia tantum, or marked for definiteness: with a definite article (English, French, Greek), with a definite article attached to the noun trigger with no pre-nominal modifiers or to the leftmost modifier in Bulgarian, and only with the definite form of adjectives in Serbian.

Bulgarian, French, Greek, and Serbian personal, location (apart from cities in French which usually do not express gender) and organization names are marked for grammatical noun gender – masculine or feminine, in contrast to the English ones. Location and organization names in Bulgarian, Greek and Serbian can be marked for neuter, as well. In Greek and Serbian, proper names have the nominative, accusative, genitive and vocative case. In Serbian, names can also be declined in the dative, locative and instrumental, while in Bulgarian vocative is observed only with some forenames.

Syntactically, proper nouns are heads of noun phrases but show restricted combinatorial properties compared to common nouns. For the five languages discussed in this paper, the forenames can be extended with one or more (rarely more than two) proper nouns: a nickname, a patronym and/or a family name. Agreement in gender and number is observed if they are of Slavic and Greek origin. For feminine surnames of Slavic origin in Serbian, the agreement in gender is allowed but not obligatory. Bulgarian, French, Greek and Serbian adjectives and Bulgarian, French and Serbian possessive pronouns change to agree in gender and number with the nouns they modify. Greek and Serbian adjectives and possessive pronouns agree in case with the head noun.

Compared to personal names, location names have a more diverse structure, while organization names show the highest complexity. Both location and orga-

nization names can be proper nouns or proper names, comprising proper and common nouns or noun phrases, which begin to function as names of geographical locations and organizations, respectively.

## 4 Names and multiword expressions

Many names are composed of more than one word and are classified as multiword names. They can comprise two or more proper nouns (*Ray Jackendoff, Merrill Lynch*); common and proper nouns (Bulgarian: *Republika Bălgariya* ‘Republic Bulgaria’); adjectives and a proper or a common noun (*International Monetary Fund, Upper Manhattan*); abbreviations (*Financial Advisors Ltd., John Smith Jr., Miami, FL*); numerals or numbers (*the Second Generative Grammar Conference; XX Generative Linguistics Conference*); verbs and adverbs with names of products such as books, movies, songs (*Someone to Watch over Me; Killing Me Softly*), etc.

Anderson (2007) provides a detailed classification of proper names, a subset of which is relevant for our study, as follows: simple opaque names (*John*); simple names that have a resemblance to a common word (*Prudence*); names based on other names (*Lincoln* – for a boulevard); names overtly derived from other names (Slavic family names); names based on compounds, some of them containing a name (*Queensland, Newtown*); names based on longer phrases – they may include another name (*the University of Queensland*) or not (*Long Island, Hen and Chicken Island*); and names based on sentences (as with titles of movies).

An important feature that systematically distinguishes location and organization names from personal names is that the TRIGGERS may be their integral part, constituting a MWE (Bulgarian: *Černo MORE* ‘Black SEA’,<sup>1</sup> English: *PRESIDENT Roosevelt Boulevard, First Investment BANK*, French: *BANQUE de France* ‘BANK of France’, Greek: *Τράπεζα της Ελλάδος* ‘BANK of Greece’, Serbian: *Jadransko MORE* ‘Adriatic SEA’, *Međunarodni SUD pravde* ‘International COURT of Justice’).

Carroll (1985) describes the non-classifying part of the location name as a name-stem (e. g., *Trafalgar* in *Trafalgar Square*) and explores rules according to which the name-stem can be used to stand for the whole name. Not only for some location names, but also for some organization names with internal triggers, the name-stem can replace the whole name, e.g., French: *LA MAISON D’ÉDITION Ha-*

<sup>1</sup>The translations in the paper are closer to literal translations than to proper ones, e.g., Bulgarian: *Republika Bălgariya* ‘Republic Bulgaria’ instead of ‘Bulgarian Republic’ or ‘The Republic of Bulgaria’, but Greek: ο Έλληνας Πρωθυπουργός ‘the Greek PRIME MINISTER’ (instead of ‘premier’, for example).

*chette* ‘the Hachette PUBLISHING COMPANY’ or just *Hachette*, Greek: η Ολυμπιακή ‘the Olympic’ for Olympic AIRWAYS.

Further, a location name may feature a **personal name** specified by a PERSONAL TRIGGER (*SAN **Jorge** River*) that cannot be omitted without loss of the name function; similarly, an organization name may feature a **personal name** specified by a PERSONAL TRIGGER (*SAN **Jose** State University*) or a **location name** specified by a LOCATION TRIGGER (***Los Angeles** CITY College*).

For the purposes of our study, we differentiate the names on the basis of their structure: i) whether the name is a MWE or a noun; ii) whether the multiword name obligatorily incorporates a trigger (an internal trigger); and iii) whether the name (either single or a MWE) is optionally specified by a trigger (an external trigger). The external triggers may be explicit or implicit, depending on the context (the City of New York, New York City vs. New York):

- Single-word personal name (*Arthur*).
- Multiword personal name (*Arthur Conan Doyle*).
- Multiword personal name, which incorporates an internal personal trigger. When people are famous, combinations with triggers such as holy, aristocratic and religious titles can be widely used and are stable (*POPE **John** Paul II*).
- Single-word personal name; it is specified by an external personal trigger (*UNCLE **John***). Kinship terms are usually combined with a single-word personal name.
- Multiword personal name; it is specified by an external personal trigger (*PROFESSOR **Steven** Pinker*).
- Single-word location name (it can coincide or not with a personal name) (*Danube, Washington*).
- Multiword location name (it may - partially - coincide or not with a personal name) (*Little Rock, SAN **Antonio***).
- Multiword location name, comprising an internal location trigger (*Rocky MOUNTAINS*). No additional location trigger of the same type can be added; being part of the name, the trigger cannot be omitted either. A multiword location name may include a personal name (and, rarely, an organization name) (*Cristina FORT*).

- Single-word location name; it is specified by an external location trigger (*RIVER Nile*).
- Multiword location name; it is specified by an external location trigger (*VOLCANO Klyučevskaya Sopka*).
- Single-word organization name (it may coincide or not with a personal name or a location name) (*Matalan, Poundland*).
- Multiword organization name (it may (partially) coincide or not with a personal or a location name) (*Mercedes Benz*).
- Multiword organization name, comprising an internal organization trigger as an integral part of the proper name. Another organization trigger of the same type cannot be added. The trigger, which is part of the name, cannot be omitted either. The multiword organization name may include a personal or a location name (*PRINCESS Basma Youth Resource CENTER, Melbourne Grammar SCHOOL*).
- Single-word organization name; it is specified by an external organization trigger (*SUPERMARKET Galaxy*).
- Multiword organization name; it is specified by an external organization trigger (*the COMPANY Business Models INC.*)

## 5 Semantic patterns for persons, locations and organizations

Names can be grouped into different semantic classes and subclasses with respect to the properties of their referents (explicated by triggers). A name from a given class (personal, location or organization) selects triggers from a particular set of semantic subclasses. For example, complex personal names are combined with triggers that define a legislative job title, executive job title, judicial position, academic position, academic title, military rank, and profession. The permissible combinations between types of names (proper nouns, MWEs), and semantic subclasses of triggers determine the semantic patterns applicable to the personal, location and organization names. The semantic patterns we propose show semantic compatibility valid for a particular semantic class and describe the permissible

combinatory options. For example, a personal name can be extended with a kinship term (i.e., *the beautiful STEP-DAUGHTER of John from Paris, Anne Nicole*) and the kinship term can be specified in various ways and restricted for possessor and location, thus the respective semantic pattern is: (modifier: referent specification phrase) – trigger: kinship term – (complement: possessor phrase) – (complement: location phrase) – personal name.

As triggers refer to concepts, the semantic relations in which they are involved should be universal and must hold among the relevant concepts in any language. Thus, the semantic patterns describe language-neutral relations and can be regarded as universal structures with correlating language-specific syntactic patterns.

Following the detailed hierarchy employed by Giuliano (2009) for automatic classification of personal NEs, we can conclude that every common noun that determines the referent of a personal name can be a trigger, i.e., words such as *chess-player, singer, footballer*, etc. Magnini et al. (2002) use WordNet hierarchy for identification of large sets of triggers – hyponyms of high-level synsets such as {person}, {location}, {organization}. Some authors suggest verb triggers appearing in the local context of NEs (Zhang et al. 2004), e.g., for water bodies (like rivers) the verb *flooded* in *The Sava flooded the village* indicates that *Sava* is a river and not a person. There are detailed classifications of NEs (of more than 200 categories; cf. Sekine & Nobata 2004), while other classifications build shallow hierarchies with the major classes on the top and sets of subtypes with different granularity at the low levels (ACE 2018; Fleischman & Hovy 2002).

In our study, we distinguish the following semantic subclasses for person, location and organization names and their triggers:

- Persons and personal triggers –legislative job title: *prime minister*; executive job title: *executive officer*; judicial position: *judge*; academic position: *associate professor*; military rank: *major general*; profession: *engineer*; academic title: *Ph.D.*; true honorific: *Mister / Mr.*; aristocratic title: *Prince*; religious title: *Bishop*; kinship term: *sister*; holy title: *Saint*.
- Locations and location triggers –natural: *river*; public: *monument*; commercial: *restaurant*; infrastructure: *boulevard*.
- Organizations and organization triggers –business: *company*; political: *political party*; government: *ministry*; media: *publishing house*; human / non-government: *association*.

We classify proper names (persons, locations and organizations) in the patterns (A) to (I) below according to their shared features. Patterns are described in terms of the categories (a)-(d): (a) the semantic subclass of the trigger; (b) the type of the proper name that selects triggers; (c) obligatoriness / optionality of the trigger manifested by an internal or external trigger with respect to the name; (d) the semantic pattern that the proper name evokes.

### 5.1 Pattern A

(a) Semantic class of the trigger: legislative job title, executive job title, judicial position, academic position, academic title, military rank, profession. Specification of military ranks and top-level legislative, executive, and judicial triggers is not allowed: \*prime minister of finance; lower level legislative triggers can be specified: engineer in automatics. (b) Type of the proper name: personal name extended or substituted by a family name. (c) External trigger. (d) Semantic pattern: (referent specification phrase) – trigger – (domain specification phrase) – (possessor phrase) – (affiliation phrase) – (location phrase).

Example: English: (*his* | *Stefan's*) (*new*) *PROFESSOR* (*of law*) (*at the University*) (*in Plovdiv*) ***Ivan Ivanov***.

### 5.2 Pattern B

(a) Semantic class of the trigger: aristocratic title, religious title. (b) Type of the proper name: personal name or family name. Some aristocratic and religious titles are selected only by a personal name (*POPE Francis*), while others are selected by a family name (*LORD Orsini*). A trigger can also be part of a personal name (for distinguished persons) but no separate pattern is defined for this type of name. (c) External trigger. (d) Semantic pattern: (referent specification phrase) – TRIGGER – (affiliation phrase) – (location phrase).

Example: English: *the* (*new*) *METROPOLITAN* (*of the Church*) (*in San Francisco*) ***Iona***.

### 5.3 Pattern C

(a) Semantic class of the trigger: kinship term. (b) Type of the proper name: personal name (rarely modified by family name(s)). (c) External trigger. (d) Semantic pattern: (referent specification phrase) – TRIGGER – (possessor phrase) – (location phrase).

Example: English: (*his* | *Ivan's*) (*blond*) *STEP-BROTHER* (*from Sofia*) ***Stefan***.

#### 5.4 Pattern D

(a) Semantic class of the trigger: holy title (a limited set of words). (b) Type of the proper name: personal name (rarely modified or substituted by a nickname). (c) External trigger. (d) Semantic pattern: (referent specification phrase) – TRIGGER – (location phrase).

Example: English: (*miraculous*) SAINT (*from Patara*) **Nicholas**.

#### 5.5 Pattern E

(a) Semantic class of the trigger: true honorific. (b) Type of the proper name: personal name extended or substituted by a family name. (c) External trigger. (d) Semantic pattern: – trigger

Example: English: **MONSIEUR Ivan Ivanov**.

#### 5.6 Pattern F

(a) Semantic class of the trigger: location. (b) Type of the proper name: location name. (c) External trigger. (d) Semantic pattern: (referent specification phrase) – TRIGGER – (specification phrase) – (possessor phrase) – (location phrase)

Example: English: *the (beautiful)* CITY (*near the big river*), **Plovdiv**.

#### 5.7 Pattern G

(a) Semantic class of the trigger: location. (b) Type of the proper name: location name. (c) Internal trigger. (d) Semantic pattern: (referent specification phrase) – internal TRIGGER – (location phrase)

Example: English: *the (beautiful)* MOUNT **Fuji** (*in Japan*).

#### 5.8 Pattern H

(a) Semantic class of the trigger: organization. (b) Type of the proper name: organization name. (c) External trigger. (d) Semantic pattern: (referent specification phrase) – TRIGGER – (domain specification phrase) – (possessor phrase) – (affiliation phrase) – (location phrase).

Example: English: *the (new)* COMPANY (*of his friends*) (*in Athens*), **Tetacom**.

## 5.9 Pattern I

(a) Semantic class of the trigger: organization. (b) Type of the proper name: organization name. (c) Internal trigger. (d) Semantic pattern: (referent specification phrase) – internal TRIGGER – (location phrase).

Example: English: *the (new) Hebros BANK (in Athens)*.

## 6 Language-specific syntactic patterns for persons, locations and organizations

We define the semantic patterns evoked by different types of proper names when combined with triggers, and the syntactic patterns that involve combinations of: modifiers, one or several, semantically restricted by the head proper noun and the trigger, and complements, semantically restricted by the trigger.

The syntactic patterns are language-specific and differ for personal, location and organization names. The syntactic patterns may involve combinations of adjectival modifiers in pre- or post-nominal position, one or several; pronoun modifiers in pre-nominal position (possessive and demonstrative); complements in post-nominal position, one or several; and a noun modifier in pre- or post-nominal position, alternating with a prepositional phrase.<sup>2</sup> Adjectival modifiers (that in Bulgarian, French, Greek and Serbian agree with the head noun in gender and number) may indicate physical shape, status, etc. Complements may indicate (domain) specification, affiliation, location, possessor, and may be prepositional or case complements, depending on the language structure.

Multiword names have the structure of a noun phrase and exhibit specific properties with respect to constituency of the head noun and the components, including various constraints on modifiers, complements, clitics (in Bulgarian and Greek), etc.

The syntactic patterns represent language-specific grammatical features and dependencies and how these features and dependencies are manifested in a particular language. One or more syntactic patterns from one or different languages may correspond to the same semantic pattern. The syntactic patterns, as they are presented in this paper, define constituency and reflect the morphological and syntactic structure of a particular language, although they do not strictly describe phrase structure and grammatical dependencies. However, the syntactic

---

<sup>2</sup>The *noun modifier – prepositional phrase* alternation is not described in the syntactic patterns.

patterns are formal enough to code the linguistic information correctly and to allow for the conversion to some formalism.

Syntactic patterns corresponding to the largely universal semantic patterns, described in §5, are formulated for English, Bulgarian, French, Greek, and Serbian. The generalizations for semantic patterns and respective syntactic patterns were constructed on the basis of observations and classifications made on dictionaries of NEs, annotated corpora of NEs and grammars for NE recognition developed so far (Krstev et al. 2013; Koeva & Dimitrova 2015).

### 6.1 Syntactic pattern A (single family name or multiword personal name)

Characteristics shared by the five languages<sup>3</sup>: i) Triggers are placed to the left of the personal name; a complex trigger phrase is likely to be an apposition<sup>4</sup>. Examples: English: *the new PROFESSOR of Sustainable Agriculture and Climate Change, Chris Smith* or *Chris Smith, the new PROFESSOR of Law at the University's Humanities Institute*; French: *Le nouveau PROFESSEUR de morale de l'Université de Fribourg, Thierry Collaud* 'the new PROFESSOR in ethics at the University of Fribourg, Thierry Collaud' or *Thierry Collaud, le nouveau PROFESSEUR de morale de l'Université de Fribourg* 'Thierry Collaud, the new PROFESSOR in ethics at the University of Fribourg'; ii) If no modifiers or complements exist, the trigger is indefinite (except for Greek where the article is obligatory); otherwise, it is definite. Examples: Bulgarian: *MINISTĀR Kuneva* 'MINISTER Kuneva' versus *noviyat MINISTĀR v pravitelstvoto Ivan Dimitrov* 'new-the MINISTER in government-the Ivan Dimitrov'; English: *PROFESSOR Chomsky* versus *the PROFESSOR of linguistics Noam Chomsky*; iii) personal names can be extended with one or more (rarely more than two) proper names: a nickname, a patronym and/or a family name, constituting a MWE. Example: Serbian: *DR Slavica Đukić-Dejanović, PREDSEDNIK Narodne skupštine* 'DR. Slavica Đukić-Dejanović, PRESIDENT of National Assembly'.<sup>5</sup>

<sup>3</sup>Language specific characteristics were also formulated but due to limitation of space they are represented only in the syntactic patterns.

<sup>4</sup>The restrictive apposition of a proper noun (whose omission changes the meaning of the sentence) is covered by the syntactic patterns (*the new PROFESSOR of Law, Chris Smith*). The non-restrictive apposition (*Chris Smith, the new PROFESSOR of Law*) is not included due to limitation of space. In this paper, the term apposition is used for the "non-restrictive apposition" only.

<sup>5</sup>The vertical bar – |, separates alternatives. The question mark indicates zero or one occurrences of the preceding element. The asterisk "\*", indicates zero or more occurrences of the preceding element. Parentheses "()", are used to define the scope and precedence of the operators. The equality sign =, indicates the semantics of the prepositions and case phrases.

**English:** (((DefArt|GenDet|GPossPron) Adj\* TRIGGER (PP in|of|for = DomSpec)? (PP at = Aff)\* (PP in = Loc)?) | (DefArt TRIGGER (PP in|of|for = DomSpec)? (PP at|of = Aff)\* (PP in = Loc)?) | INDEFTRIGGER) **PerN**

**Bulgarian:** (((DefAdj (PossCL)?) | DefPossPron) Adj\* TRIGGER (PP po|na|za = DomSpec)? (PP na|pri|v|kăm|ot = Aff)\* (PP ot = Loc)?) | (DEFTRIGGER (PossCL)? (PP po|na|za = DomSpec)? (PP na|pri|v|kăm|ot = Aff)\* (PP ot = Loc)?) | INDEFTRIGGER) **PerN**

**French:** (((DefArt|PossPron) Adj\* TRIGGER (PP en|de = DomSpec)? (PP à = Aff)\* (PP à = Loc)?) | (DefArt TRIGGER Adj\* (PP en|de = DomSpec)? (PP à|de = Aff)\* (PP à = Loc)?) | INDEFTRIGGER) **PerN**

**Greek:** ((DefArt Adj\* (PossPron)? TRIGGER (GenP = DomSpec)? (PP σε = Aff)\* (PP σε = Loc)?) | (DefArt TRIGGER (GenP = DomSpec)? (GenP = Aff)\* (PP σε = Loc)?) | DEFTRIGGER) **PerN**

**Serbian:** ((Adv | PossPron)? DefAdj\* TRIGGER (PP za| GenP = DomSpec)? (PP u|pri | GenP = Aff)\* (PP u|za|iz GenP = Loc)?) **PerN**

Table 1: Syntactic pattern A – an example translated in the five languages. The examples do not illustrate all variants.

English	Spain’s SECRETARY of State for Foreign Affairs, <b>Gonzalo de Benito Secades</b>
Bulgarian	Dържавният SEKRETAR na vãnšnite raboti na Ispaniya, <b>Gonzalo de Benito Sekades</b>
French	le MINISTRE espagnol des Affaires Etrangères, <b>Gonzalo de Benito Secades</b>
Greek	ο Ισπανός Υπουργός Εξωτερικών, <b>Gonzalo de Benito Secades</b>
Serbian	Državni SEKRETAR za spoljne poslove Španije, <b>Gonzalo de Benito Sekades</b>

## 6.2 Syntactic pattern B (single- or multiword personal name)

Characteristics shared by the five languages: i) The trigger phrase is placed in front of the personal name but the appositive order can also be found, especially if the trigger phrase is complex. Examples: English: *Venerable Dionysius, the ARCHIMANDRITE of St Sergius’ Monastery*; Serbian: *njegovo preosveštenstvo EPISKOP*

*niški Irinej* ‘His Grace BISHOP of-Niš **Irinej**’; ii) If no trigger modifiers or complements exist, the trigger is indefinite (except for Greek where the article is obligatory); otherwise, it is definite. Examples: Bulgarian: *PATRIARH Maksim* ‘Patriarch **Maxim**’; French: *L’ARCHEVÊQUE de Paris, Monseigneur André Vingt-Trois* ‘the ARCHBISHOP of Paris, Monsignor **André Vingt-Trois**’; *Le bienheureux PÈRE Brottier* ‘the blessed FATHER **Brottier**’; iii) personal names can be extended with one or more (rarely more than two) proper names: a nickname, a patronym and/or a family name. These constituents form a complex name (MWE). Example: Greek: ο πρόσφατα χειροτονηθείς Σεβασμιότατος Μητροπολίτης Κεφαλληνίας, πατέρας Γεώργιος Σαπουνάς ‘the newly appointed Most Reverend BISHOP METROPOLITAN of Kefalonia, FATHER **Georgios Sapounas**’.

**English:** ((DefArt Adj\* TRIGGER (PP of = Aff)\* (PP in = Loc)?) | (DefArt TRIGGER (PP at/of = Aff)\* (PP in = Loc)?) | INDEFTRIGGER) PerN

**Bulgarian:** (((DefAdj (PossCL)? | DefPossPron) Adj\* TRIGGER (PP na|pri|v|kăm|ot = Aff)? (PP ot = Loc)?) | (DEFTRIGGER (PP na|pri|v|kăm|ot = Aff)? (PP ot = Loc)?) | INDEFTRIGGER) PerN

**French:** ((DefArt Adj\* TRIGGER (PP de = Aff)\* (PP à|en = Loc)?) | (DefArt TRIGGER\* Adj\* (PP de = Aff)\* (PP à|en = Loc)?) | INDEFTRIGGER) PerN

**Greek:** ((DefArt Adj\* TRIGGER (GenP = Aff)\* (GenP = Loc)?) | (DefArt TRIGGER (GenP = Aff | Loc)\* (PP σε = Loc)?) | DEFTRIGGER) PerN

**Serbian:** (((PossPron | PossAdj) TRIGGER) | (DefAdj\* TRIGGER) | (TRIGGER Def-Adj?)) (PP u|za| GenP = Aff)? (PP u|za|iz GenP = Loc)? PerN DefAdj\*

Table 2: Syntactic pattern B – an example translated in the five languages.

English	<i>the new BISHOP of the Christian Catholic Church of Switzerland, DR <b>Harald Rein</b></i>
Bulgarian	<i>noviyat EPISKOP na Hristiyanskata katoličeska cърkva v Šveycariya, D-R <b>Harald Rein</b></i>
French	<i>le nouvel EVÊQUE de l’Eglise catholique-chrétienne de la Suisse, DR <b>Harald Rein</b></i>
Greek	<i>ο νέος Αρχιεπίσκοπος της Χριστιανικής Καθολικής Εκκλησίας της Ελβετίας, Δρ. <b>Harald Rein</b></i>
Serbian	<i>novoiimenovani BISKUP Starokatoličke crkve Švajcarske, DR <b>Harald Rajn</b></i>

## 6.3 Syntactic pattern C (single- or, rarely, multiword personal name)

Characteristics shared by the five languages: i) The usual position of the trigger is in front of the personal name. The reverse order indicates apposition. Examples: Bulgarian: *moyata hubava SESTRA Ana* ‘my-the beautiful SISTER Anna’; Greek: η αδερφή του Πέτρου, η **Μαρία** ‘the SISTER of Peter, the **Maria**’; η **Μαρία**, η αδερφή του Πέτρου ‘the **Maria**, the SISTER of Peter’. ii) The trigger is accompanied by modifiers or complements. Examples: English: *his older STEP-BROTHER from Paris, Stefan*; Serbian: *njegov rođeni BRAT četvorogodišnji Zoran* ‘his birth BROTHER 4-year-old **Zoran**’; iii) The phrase headed by the trigger is definite. Examples: French: *la SŒUR de Marc, Marie* ‘the SISTER of Marc, **Maria**’ vs. *sa SŒUR, Marie* ‘his SISTER, **Maria**’; *son BEAU-FRÈRE de Paris, Jean* ‘his BROTHER-IN-LAW from Paris, **John**’.

**English:** (((DefArt | GenDet | PossPron) Adj\* TRIGGER (PP *from* = Loc)?) | (DefArt Adj\* TRIGGER (PP *of* = Poss<sup>6</sup>)? (PP *from* = Loc)?) ) PerN

**Bulgarian:** (((DefAdj (PossCL)?) | DefPossPron) Adj\* TRIGGER (PP *ot* = Loc)?) | ((DEFTRIGGER (PossCL)?) (PP *ot* = Loc)?) | (DefAdj Adj\* TRIGGER (PP *na* = Poss)?) | (DEFTRIGGER (PP *na* = Poss)?) ) PerN

**French:** (DefArt | PossPron) Adj\* TRIGGER ((PP *de* = PerN) | (PP *de* = Loc)?) PerN

**Greek:** (DefArt Adj\* TRIGGER ((GenP = PerN) | (GenP = Poss)?) (PP *από* DefArt = Loc)?) DefArt PerN

**Serbian:** (PerN\* TRIGGER\* (DefAdj\* | (GenP = PerN))?) (PP *iz* = Loc)?) PerN | (PossPron? DefAdj\* TRIGGER\* (PP *iz* = Loc)?) PerN

Table 3: Syntactic pattern C – an example translated in the five languages.

English	his beautiful SISTER-IN-LAW from Athens, <b>Maria</b>
Bulgarian	negovata krasiva SNAHA ot Atina, <b>Maria</b>
French	sa jolie BELLE-SOEUR d’Athènes, <b>Maria</b>
Greek	η όμορφη κουνιάδα του από την Αθήνα, η <b>Μαρία</b>
Serbian	njegova lepa SNAJA iz Atine, <b>Marija</b>

<sup>6</sup>The possessor phrase is shown only for the kinship terms.

#### 6.4 Syntactic pattern D (single- and, rarely, multiword personal name)

Characteristics shared by the five languages: i) The trigger appears before the personal name but a complex trigger phrase often occurs in apposition. Examples: English: *SAIN*T **Haralambos**, *the Holy MARTYR of Magnesia*; Serbian: *SVETI mučenik i arhidakon Lavrentije* ‘*SAIN*T martyr and archdeacon **Lavrentije**’. ii) If no modifiers or complements exist, the trigger is indefinite (except for Greek where the article is obligatory); otherwise, it is definite. Examples: Bulgarian: *SVETI Nikola* ‘*SAIN*T Nicholas’; French: *le SAIN*T *de l’Arcadie: Charles De Menou D’Aulnay* ‘the *SAIN*T of Arcadia: **Charles De Menou D’Aulnay**’.

**English:** (DefArt Adj\* TRIGGER (PP *of|from = Loc*)? PerN) | (DefArt Adj\* TRIGGER PerN (PP *of|from = Loc*)?) | (TRIGGER PerN (PP *of|from = Loc*)?)

**Bulgarian:** (((DefAdj (PossCL)? | DefPossPron) Adj\* TRIGGER (PP *na|ot = Loc*)?) | (DefAdj Adj\* TRIGGER (PP *na|ot = Loc*)?) | (DEFTRIGGER (PP *na|ot = Loc*)?) | INDEFTRIGGER) PerN

**French:** (DefArt Adj\* TRIGGER (PP *de = Loc*)? PerN) | (TRIGGER Adj\* PerN (PP *de = Loc*)?)

**Greek:** (DefArt Adj\* TRIGGER (GenP = Loc)? PerN) | (DefArt TRIGGER PerN (GenP = Loc)?)

**Serbian:** (PossPron? DefAdj\* TRIGGER ((PP *iz|u|sa = Loc*)? DefAdj? (N)?) PerN PossAdj? | PerN PossAdj? (PP *iz|u|sa = Loc*)?)

Table 4: Syntactic pattern D – an example translated in the five languages.

<i>English</i>	<i>the Holy MARTYR Chrysostomos of Smyrna</i>
<i>Bulgarian</i>	<i>Svetiyat MAČENIK Hrisostom ot Smirna</i>
<i>French</i>	<i>le Saint MARTYR Chryssostomos de Smyrne</i>
<i>Greek</i>	<i>ο Άγιος Ιερομάρτυρας Χρυσόστομος Σμύρνης</i>
<i>Serbian</i>	<i>Sveti MUČENIK Hrizostom Smirnski</i>

### 6.5 Syntactic pattern E (family name or multiword personal name)

Characteristics shared by the five languages: i) The external trigger usually appears before the personal name, except for some triggers that can also appear after the personal name. Examples: English: *MR. Smith*; Serbian: *Dušan Rašković*, *DIPL. OEC.*; ii) No modifiers and/or complements are allowed. Example: French: *M. Dupont* ‘MR. Dupont’; iii) If the trigger is indefinite, an article is not permissible (except for Greek where the article is obligatory). Example: English: \*(*the*) *DR. Livingstone*; iv) personal names can be extended with one or more (rarely more than two) proper names: a nickname, a patronym and/or a family name that form a complex name (MWE). Example: Bulgarian: *GOSPODIN Ivan Ivanov* ‘MISTER Ivan Ivanov’.

**English:** Bulgarian: French: INDEFTRIGGER PerN

**Greek:** DefArt TRIGGER PerN

**Serbian:** (TRIGGER PerN) | (PerN TRIGGER)

Table 5: Syntactic pattern E – an example translated in the five languages.

<i>English</i>	<i>DR. Mary Andrew Smith</i>
<i>Bulgarian</i>	<i>D-R Meri Andryu Smit</i>
<i>French</i>	<i>DR Mary Andrew Smith</i>
<i>Greek</i>	<i>η Δρ. Μαίρη Άντρίου Σμιθ</i>
<i>Serbian</i>	<i>DR Meri Endru Smit</i>

### 6.6 Syntactic pattern F (single- or multiword location name)

Characteristics shared by the five languages: i) The trigger appears before the personal name. A heavy trigger phrase is often found in apposition. Examples: Greek: *η όμορφη πόλη των Παρισίων / του Παρισιού* ‘the beautiful CITY of Paris’; *η Σαντορίνη*, *το πιο όμορφο νησί της Ελλάδας* ‘the Santorini, the most beautiful ISLAND in Greece’. ii) The phrase headed by the trigger is definite. Examples: English: *the beautiful CITY of artists, Plovdiv*; *our beautiful CITY, Plovdiv*; French: *la belle VILLE des mille fontaines, Aix-en-Provence* ‘the beautiful CITY of thousand fountains, Aix-en-Provence’; iii) Location names can be MWEs. Examples: Bulgarian: *HRAM-PAMETNIK Sveti Aleksandăr Nevski* ‘CATHEDRAL Saint Alexander Nevski’; Serbian: *NADOŠLI Beli Timok* ‘RISING Beli Timok’.

**English:** (((DefArt | GenDet | PossPron) Adj\* TRIGGER (PP *of* = *Spec*)? (PP *in* = *Loc*)?) | (DefArt TRIGGER (PP *of* = *Spec*)?) | INDEFTRIGGER) **LocN**

**Bulgarian:** (((DefAdj (PossCL)? | DefPossPron) Adj\* TRIGGER (PP *na* = *Spec*)? (PP *v|na|pri* = *Loc*)?) | (DEFTRIGGER (PP *na* = *Spec*)? (PP *v|na|pri* = *Loc*)?) | INDEFTRIGGER) **LocN**

**French:** (((DefArt | PossPron) Adj\* TRIGGER (PP *de* = *Spec*)? (PP *de* = *Loc*)?) | (DefArt TRIGGER Adj\* (PP *de* = *Spec*)?)) **LocN**

**Greek:** (((DefArt | PossPron) Adj\* TRIGGER (GenP = *Spec*)? (PP *σε* = *Loc*)?) | (DefArt TRIGGER (GenP)?)) **LocN**

**Serbian:** ((PossPron | PossAdj)? DefAdj\* TRIGGER (PP *u|na|pri|pod|u blizini* | *GenP* = *Loc*)?) **LocN**

Table 6: Syntactic pattern F – an example translated in the five languages.

English	the most romantic CITY in the world, <b>Paris</b>
Bulgarian	nay-romantičnijat GRAD v sveta, <b>Pariž</b>
French	la plus romantique VILLE dans le monde, <b>Paris</b>
Greek	η πιο ρομαντική πόλη του κόσμου, το <b>Παρίσι</b>
Serbian	najromantičniji GRAD na svetu, <b>Pariz</b>

## 6.7 Syntactic pattern G (multiword location name)

Characteristics shared by the five languages: i) The internal trigger is part of the location name, thus the location name is always a MWE. Examples: Bulgarian: *našiyat hubav GRAD Novi han* ‘our-the beautiful CITY **Novi han**’; Greek: *ο Ινδικός Ωκεανός* ‘the **Indian OCEAN**’; ii) A location name with an internal trigger is fixed, the order of constituents cannot be changed and insertions are not allowed. Examples: Bulgarian: *našata Stara PLANINA* ‘our-the **Stara PLANINA**’, \**PLANINA stara*; French: *le célèbre MONT Blanc* ‘the famous **MONT Blanc**’, \**Blanc MONT*. iii) The location name may contain a personal or a location name (rarely an organization name). Example: English: *Minnesota RIVER*; iv) The internal trigger may be specified by the same range of modifiers and complements

permissible for the trigger.<sup>7</sup> Example: Bulgarian: **Černi vrāh** ‘Black PEAK’; v) External trigger can be added if different from the internal one. Example: English: **CITY of Colorado SPRINGS**; vi) Heavy trigger phrases often occur as appositions. Example: Serbian: **Novo Brdo, najveći rudarski GRAD u Srbiji i na celom Balkan- skom poluostrvu** ‘New Hill, biggest mining TOWN in Serbia and on the entire Balkan Peninsula’.

**English:** ((DefArt | GenDet | PossPron) Adj\* MWLocN (PP *in = Loc*)?) | (DefArt MWLocN (PP *in = Loc*)?) | MWLocN

**Bulgarian:** (((DefAdj (PossCL)?) | DefPossPron) Adj\* MWLocN (PP *v|na|pri = Loc*)?) | (DefMWLocN (PP *v|na|pri = Loc*)?) | MWLocN

**French:** ((DefArt | PossPron) Adj\* MWLocN (PP *de = Loc*)?) | (MWLocN (PP *de = Loc*)?) | MWLocN

**Greek:** ((DefArt | PossPron) Adj\* MWLocN (GenP = *Loc*)?) | (DefMWLocN (GenP = *Loc*)?) | MWLocN

**Serbian:** ((PossPron | PossAdj)? DefAdj\* (PP *u|na|pri|pod|u blizini* | GenP = *Loc*)?) MWLocN | GenP = *Loc*)?) | MWLocN

Table 7: Syntactic pattern G – an example translated in all five languages.

English	the vast <b>Great PLAINS</b> in the United States
Bulgarian	neobyatnite <b>Golemi RAVNINI</b> v Săedinenite štati
French	les <b>Grandes PLAINES</b> aux Etats-Unis
Greek	τα <b>Great PLAINS</b> της Αμερικής
Serbian	prostrane <b>Velike RAVNICE</b> u Sjedinjenim Državama

## 6.8 Syntactic pattern H (single- and multiword organization name)

Characteristics shared by the five languages: i) The trigger may either be placed before or after the organization name in case of apposition. Examples: Bulgarian: *hranitelnata KOMPANIYA* “**Danon**” ‘nutritional-the COMPANY **Danone**’; English:

<sup>7</sup>For simplicity, not all variants are presented in the syntactic pattern (applicable also to Syntactic pattern I in §6.9).

**Apache CORP.** ii) The phrase headed by the trigger is definite. If the trigger is a single-word one or specified for domain, the trigger phrase may be indefinite (except for Greek where the article is obligatory). Examples: English: *the COMPANY of Buffett, Berkshire Hathaway*; Bulgarian: *investicionen FOND “Razvitie”* ‘investment FUND Razvitie’; French: *la nouvelle COMPAGNIE Santus* ‘the new COMPANY Santus’, *notre nouvelle COMPAGNIE Santus* ‘our new COMPANY Santus’, *la nouvelle COMPAGNIE de Pierre, Santus* ‘the new COMPANY of Pierre, Santus’; iii) The organization name can be a MWE – either a complex personal or location name or a fixed multiword organization name. Examples: Bulgarian: *novoto UČILIŠTE za deca s uvreden sluh v Sofiya “Prof. Dr. Dečo Denev”* ‘new-the school for children with impaired hearing in Sofia Prof. Dr. Dečo Denev’; Greek: *η μεταλλευτική εταιρία Ελληνικός Χρυσός* ‘the mining COMPANY Hellas Gold’, *η Ελληνικός Χρυσός, μεταλλευτική εταιρία* ‘the Hellas Gold mining COMPANY’; Serbian: *FABRIKA mašina “Ivo Lola Ribar”* ‘Machine FACTORY “Ivo Lola Ribar”’.

**English:** (((DefArt | GenDet | PossPron) Adj\* TRIGGER (PP *in|at* = *DomSpec*)? (PP *inat* = *Loc*)?) | (DefArt Adj\* TRIGGER (PP *inat* = *DomSpec*)? (PP *in|at* = *Loc*)?) | TRIGGER) **OrgN**

**Bulgarian:** (((DefAdj (PossCL)?) | DefPossPron) Adj\* TRIGGER (PP *po|na|za* = *DomSpec*)? (PP *v|na|pri* = *Loc*)?) | ((DEFTRIGGER (PossCL)?) (PP *po|na|za* = *DomSpec*)? (PP *v|na|pri* = *Loc*)?) | TRIGGER) **OrgN**

**French:** (((DefArt | PossPron) Adj\* TRIGGER (PP *de* = *DomSpec*)?) | (DefArt Adj\* TRIGGER (PP *de* = *PersN*)?) | (DefArt TRIGGER)) (PP *σε* DefArt = **OrgN**)

**Greek:** ((DefArt Adj\* TRIGGER (PossPron)?) ((PP *για* = *DomSpec*)?) | (*Gen* = *DomSpec*)?) | (DefArt Adj\* TRIGGER (GenDet)?) | (DefArt TRIGGER)) (PP *σε* DefArt = **OrgN**)

**Serbian:** (PossPron? DefAdj\* TRIGGER (PP *za* | *GenP* = *DomSpec*)? (PP *u|pri* | *GenP* = *Aff*)? (PP *iz|u* | *GenP* = *Loc*)? **OrgN**) | (**OrgN** TRIGGER)

Table 8: Syntactic pattern H – an example translated in the five languages.

English	China’s investment BANK, <b>China International Capital Corporation Limited</b>
Bulgarian	<b>kitayskata</b> investicionna BANKA <b>China International Capital Corporation Limited</b>
French	la BANQUE d’Investissements en Chine, <b>China International Capital Corporation Limited</b>
Greek	η Τράπεζα Επενδύσεων της Κίνας, η <b>China International Capital Corporation Limited</b>
Serbian	kineska investiciona BANKA, <b>Kineska međunarodna kapitalna korporacija</b>

### 6.9 Syntactic pattern I (multiword organization name)

Characteristics shared by the five languages: i) The trigger is an integral part of the organization name, thus the organization name is always a MWE. Examples: English: *the European BANK for Reconstruction and Development in Serbia*; *the ASSOCIATION of Chartered Certified Accountants*; Greek: η Τράπεζα Εμπορίου και Ανάπτυξης της Μαύρης Θάλασσας ‘the Black Sea Trade and Development BANK’. ii) Organization names containing an integral trigger are fixed, the order of constituents cannot be changed and insertions are not allowed. Examples: Bulgarian: *Evropeyska BANKA za vǎzstanovyavane i razvǎtie* ‘European BANK for Reconstruction and Development’; *novosǎzdateniyat Evropeyski FOND za strategičeski investicii* ‘newly-found-the European FUND for Strategic Investment’; iii) organization names with an integral trigger can contain a personal, location or organization name. Example: Serbian: *Memorialni CENTAR “Josip Broz Tito”* ‘Memorial CENTER “Josip Broz Tito”’; iv) The internal organization trigger can be specified by the same range of modifiers and complements permissible for it in a regular use. Example: French: *L’ASSOCIATION des Historiens* ‘the ASSOCIATION of Historians’; v) Rarely, an organization trigger, different from the integral trigger, can specify the multiword organization name. Examples: Bulgarian: *SǎYUZ na tǎrgovcite v Bǎlgariya* ‘UNION of traders-the in Bulgaria’; *ASOCIACIYA “SǎYUZ na tǎrgovcite v Bǎlgariya”* ‘ASSOCIATION UNION of traders-the in Bulgaria’.

**English:** (DefArt | PossPron) Adj\* MWOrgN (PP at|in = Loc)? | (DefMWOrgN (PP at|in = Loc)?) | MWOrgN

**Bulgarian:** (((DefAdj (PossCL)? | DefPossPron) Adj\* MWOrgN (PP v|na|pri = Loc)?) | (DefMWOrgN (PP v|na|pri = Loc)?) | MWOrgN

**French:** (DefArt | PossPron) Adj\* MWOrgN (PP à de = Loc)?

**Greek:** ((DefArt | PossPron) Adj\* MWOrgN (PP σε = Loc)?) | (DefArt MWOrgN (GenP = Loc)?) | MWOrgN

**Serbian:** (PossPron? DefAdj\* MWOrgN (GenP = Poss)? (PP iz|u| GenP = Loc)?) | MWOrgN

Table 9: Syntactic pattern I – an example translated in all five languages.

English	<b>World ASSOCIATION for Small and Medium Enterprises</b>
Bulgarian	<b>Svetovna ASOCIACIYA za malki i sredni predpriyatiya</b>
French	<b>ASSOCIATION Mondiale des Petites et Moyennes Entreprises</b>
Greek	<b>Διεθνής Σύνδεσμος Μικρομεσαίων Επιχειρήσεων</b>
Serbian	<b>Svetsko UDRUŽENJE malih i srednjih preduzeća</b>

## 7 Comparison of the five languages

At the semantic level, languages do not differ (significantly). The semantic patterns of proper names define the common semantics, regardless of the language in which it is realized: semantic patterns are language-neutral. Languages differ in lexical and phrasal categories, constituency, word order permutations and alterations. The differences in word order and alterations insert some nuances in the expressed meaning, i.e., the viewpoint of the speaker, but they do not alter the general meaning. The syntactic patterns of proper names show the correspondences among languages at the syntactic level: syntactic patterns are language-specific. A semantic pattern is a representation that can be linked with different syntactic frames in different languages and, vice versa, syntactic patterns from different languages may share a single semantic pattern. Thus, syntactic patterns make explicit the similarities and differences in the grammatical structure of the five languages.

The structure of the language-neutral semantic and language-specific syntactic patterns can be represented as a graph whose nodes are semantic and syntactic patterns while the arcs represent different languages. More than one language-specific syntactic pattern may be linked to one language-neutral semantic frame; in such a case, syntactic patterns are synonymous to the extent that they represent a common semantic structure. Through this type of representation, we offer an interlingual mapping of the syntactic structures of named entities in the five languages. Some of the most distinctive grammatical characteristics of NEs in English, Bulgarian, French, Greek and Serbian with respect to the single and multiword morphology and syntax will be outlined below.

### 7.1 Grammatical categories of dependent constituents

The syntactic patterns of English proper name triggers involve combinations of adjectival modifiers in pre-nominal position, one or several (they can be preceded by a definite article) (*the great POET Burns*); possessive pronoun modifiers in pre-nominal position (*Welcome our new PROFESSOR, Jennifer S. Locke!*); prepositional complements in post-nominal position; a noun modifier in pre-nominal position, alternating with a genitive determiner and a prepositional phrase, e.g., *the Grieg Piano CONCERTO* vs. *Grieg's Piano CONCERTO* vs. *the Piano CONCERTO by Grieg*.

In Bulgarian, the syntactic patterns for proper name triggers exhibit the following combinations: adjectival modifiers in pre-nominal position (*noviyat PREDSEDATEL Petrov* 'new-the CHAIR Petrov'); possessive pronoun modifiers in pre-nominal position alternating with a possessive pronoun clitic in post-nominal position (*moyata sestra Ana* 'my-the SISTER Ana' vs. *SESTRA mi Ana* 'SISTER my.PossCL Ana'); prepositional complements in post-nominal position (*KOMPANIYATA na Ivan "Elit"* 'COMPANY-THE of Ivan "Elit"'); and a noun modifier in pre-nominal position, alternating with a PP (*karate INSTRUKTORĀT Ivan* 'karate TRAINER-THE Ivan' vs. *INSTRUCTORĀT po karate Ivan* 'TRAINER-THE in karate Ivan').

In French, the phrase headed by a trigger is definite (with an alternation of the phrase with a definite article or possessive pronoun: *la belle VILLE des Mille Fontaines, Aix-en-Provence* 'the beautiful CITY of thousand fountains, Aix-en-Provence'; *notre belle VILLE, Paris* 'our beautiful CITY, Paris'). The proper name can be introduced by a preposition: *la belle VILLE de Paris* 'the beautiful CITY of Paris'; *notre belle VILLE de Paris* 'our beautiful CITY of Paris'.

In Greek, simple and multiword proper names are preceded by a definite article, e.g., *το Παρίσι* 'the Paris', *οι Ηνωμένες Πολιτείες Αμερικής* 'the United

States of America’. The phrase headed by the trigger is also definite: η όμορφη πόλη των καταρρακτών, η Έδεσσα ‘the beautiful CITY of waterfalls, the Edessa’; η όμορφή μας πόλη, η Έδεσσα ‘the beautiful our.PossCL CITY, the Edessa’. A location name can be put in the genitive case: η όμορφη πόλη του Παρισιού ‘the beautiful CITY of Paris’. In that case, the use of the possessive pronoun clitic is not possible: \*η όμορφή μας πόλη του Παρισιού ‘the beautiful our.PossCL CITY of Paris’.

The syntactic patterns for Serbian names comprise combinations of adjectival modifiers in pre-nominal position (*Američka filmska* AKADEMIJA ‘American Film ACADEMY’); sometimes they can be found in post-nominal position (*EPISKOP niški Irinej* ‘BISHOP of-Niš Irinej’); prepositional complements in post-nominal position (*AZOTARA u Pančevu* ‘Fertilizer PLANT in Pančevo’); complements in genitive (alternating with a prepositional phrase) – *DOM zdravlja* ‘HOUSE (of)-Health.Gen (Community Health Center)’ and *DIREKCIJA za upravljanje oduzetom imovinom* ‘DIRECTORATE for Management of Seized Assets’.

Coordinated phrases are possible in all languages, e.g., Bulgarian: *VICEPREMIERĀT i MINISTĀR na obrazovaniето i naukata, Meglena Kuneva* ‘Deputy PRIME MINISTER -THE and MINISTER of Education-the and Science-the, Meglena Kuneva’; French: *Martin Vetterli, PROFESSEUR à l’Ecole polytechnique fédérale de Lausanne et PRÉSI-DENT du Conseil national de la recherche* ‘Martin Vetterli, PROFESSOR at the Federal Polytechnic School of Lausanne and PRESIDENT of the National Research Council’; Greek: ο πρωθυπουργός και πρόεδρος του ΣΥΡΙΖΑ Αλέξης Τσίπρας ‘the PRIME MINISTER and HEAD of Syriza Alexis Tsipras’; Serbian: *Hleb i kifle* ‘Bread and Rolls’ (an organization name). Both triggers and proper names can appear in a coordinated construction (we do not encode coordination in the syntactic patterns).

## 7.2 Definiteness

Definiteness is expressed either by a morpheme as in Bulgarian and Serbian, or by an article as in English, French, and Greek. In English, French, and Greek, the definite article precedes the trigger, e.g., *le PREMIER MINISTRE, Justin Trudeau* ‘the PRIME MINISTER, Justin Trudeau’. There are other means to express definiteness – i.e., the demonstrative pronouns, the possessive pronouns in English, French, and Serbian, e.g., French: *notre belle VILLE, Paris* ‘our beautiful CITY, Paris’; Serbian: *njeno rodno MESTO Beograd* ‘her native CITY Belgrade’.

With personal names in English, the definite article is obligatory when it is modified by an adjective and/or a PP complement: *the great POET Burns, the Scottish POET Burns; the POET from Kosovo, Fahredin Shehu; the AUTHOR of the*

*Concerto, Edvard Grieg*. The possessive pronoun, the article and the genitive determiner are in complementary distribution in English.

The definite form in Bulgarian is required when the trigger is modified by an adjective or a possessive pronoun (in this case the definite adjective is part of the first phrasal constituent: *noviyat* *MINISTĀR* *Valentin Dimitrov* ‘new-the MINISTER *Valentin Dimitrov*’, and / or a prepositional phrase; if there are no pre-nominal modifiers, the article is on the trigger word: *MINISTĀRĀT na finansite* *Valentin Dimitrov* ‘MINISTER-THE of finance-the *Valentin Dimitrov*’).

In Greek, all proper names are preceded by a definite article, e.g., η *Μαρία* ‘the *Maria*’, η *Αθήνα* ‘the *Athens*’, ο πρωθυπουργός *Αλέξης Τσίπρας* ‘the PRIME MINISTER *Alexis Tsipras*’.

In Serbian, the definite article is not used; furthermore neither possessive pronouns nor adjectives are obligatory. However, when adjectives precede proper names, they are in definite form, e.g., *od izvesnog* *Stevice Miletića* vs. \**od izvesna* *Stevice Miletića* ‘from certain *Stevica Miletić*’.

### 7.3 Distribution of clitics in Bulgarian and Greek

The possessive pronoun clitics in Bulgarian are right-adjacent to the definite article, e.g., in the second position in the noun phrase (*noviyat ni* *MINISTĀR* *Valentin Dimitrov* ‘new-the our.PossCL MINISTER *Valentin Dimitrov*’ vs. *MINISTĀRĀT ni na finansite* *Valentin Dimitrov* ‘MINISTER-THE our.PossCL of finance-the *Valentin Dimitrov*’). The possessive pronoun clitic in Bulgarian is also right-adjacent to an indefinite kinship term if used without an adjectival modifier as in *MAYKA mi* *Maria* ‘MOTHER my.PossCL *Maria*’.

In Bulgarian the interrogative particle *li* (which is always a clitic) may appear after the first definite modifier (if not followed by a possessive pronoun clitic) or after the whole NP, as in: *noviyat li* *DIREKTOR* *Ivanov* ‘new-the li.QuCL DIRECTOR *Ivanov*’ and *noviyat* *DIREKTOR* *Ivanov* *li* ‘new-the DIRECTOR *Ivanov* li.QuCL’). The above-stated rules for the definite article, possessive pronoun clitic and interrogative particle hold for the multiword names too, with the leftmost adjective being part of the proper name itself (*Bălgarskata narodna* *BANKA* ‘*Bulgarian-the National BANK*’).

Greek pronoun clitics are right-adjacent to the proper name, e.g., η *Μαρία* μου ‘the *Maria* my.PossCL’. Once there is a trigger followed by the proper name, the possessive pronoun clitic is between the trigger and the proper name, e.g., ο καθηγητής μας *Χρήστος Τσολάκης* ‘the PROFESSOR our.PossCL *Christos Tsolakis*’.

In Serbian, pronoun clitics can sometimes be used to express possession, as in *KOMŠIJA mi Asan* ‘NEIGHBOR I.CL Asan (my neighbor Asan)’. However, these constructions are rarely used, being considered rather obsolete and non-standard and are therefore not included in the patterns.

#### 7.4 Expression of semantic and grammatical dependencies

Prepositions are used to express semantic and grammatical dependencies, such as affiliation, domain specification, location in English, Bulgarian, French, Greek, and Serbian, and possession in English, Bulgarian, and French. Semantic and grammatical dependencies can be signified by cases in Greek and Serbian. In English, possession may also be expressed by a clitic *-’s* (marking the genitive determiner), and in Bulgarian by the derivational suffix of possessive adjectives.

#### 7.5 Word order – position of the trigger with respect to the proper noun

In French, Greek and Serbian, word order permutations are common for personal names, as the first name and surname(s) can change places: French: *Nicolas Sarkozy* vs. *Sarkozy Nicolas*; Greek: Γεώργιος Κοκκινόπουλος ‘Georgios Kokkinopoulos’ vs. Κοκκινόπουλος Γεώργιος ‘Kokkinopoulos Georgios’; Serbian: *Marko Vitas* vs. *Vitas Marko*. In Serbian, a change of the order of the first name and the surname(s) of male persons results in a change of the syntactic properties, as in the former case both names inflect, while in the latter only the first name inflects, e.g., in the genitive *Marka Vitasa* vs. *Vitas Marka*.

In all languages, the trigger can appear in pre- or post-nominal position: French: *le MINISTRE des Finances et des Comptes publics, Michel Sapin* ‘the MINISTER of finance and of public accounts, Michel Sapin’, or *Michel Sapin, MINISTRE des Finances et des Comptes publics* ‘Michel Sapin, MINISTER of finance and of public accounts’; Greek: ο Υπουργός Οικονομικών Ευκλείδης Τσακαλώτος ‘the MINISTER of finance Efkleidis Tsakalotos’, or ο Ευκλείδης Τσακαλώτος, Υπουργός Οικονομικών ‘the Efkleidis Tsakalotos, MINISTER of finance’. Some abbreviations can appear only before or after the names, as in: Serbian: *ЈП “Srbijašume”* ‘PC (Acronym for Public Company) Srbijašume’ but *Takovo d.o.o.* ‘Takovo (a place name) d.o.o.’.

In all languages, a complex trigger phrase is often in apposition (when the trigger appears as an apposition, it is always separated with a comma), e.g., English: *Chris, the new PROFESSOR of Agriculture and Forestry*; French: *le MINISTRE des Finances et des Comptes publics, Michel Sapin* ‘the MINISTER of finance and of pub-

lic accounts, *Michel Sapin*’, or *Michel Sapin*, *MINISTRE des Finances et des Comptes publics* ‘*Michel Sapin*, MINISTER of finance and of public accounts’; Greek: ο Καθηγητής Γεώργιος Μπαμπινιώτης ‘the PROFESSOR *Georgios Babiniotis*’ or ο Γεώργιος Μπαμπινιώτης, Καθηγητής ‘the *Georgios Babiniotis*, PROFESSOR’.

In all languages, the head personal name can be specified by more than one triggers in a preferred order of appearance: English: *DIRECTOR General PROF. Smith*; Bulgarian: *generalniyat DIREKTOR PROF. Smit* ‘general-the DIRECTOR PROF. *Smith*’; French: *DIRECTEUR général PROF. Smith* ‘DIRECTOR General PROF. *Smith*’; Greek: ο Γενικός Διευθυντής Καθηγητής Σμιθ ‘the General DIRECTOR PROFESSOR *Smith*’; Serbian: *Generalni DIREKTOR PROF. Smit* ‘General DIRECTOR PROF. *Smith*’.

## 7.6 Alternations

In English, genitive determiners may alternate with a possessive prepositional phrase.

The possessive PP in Bulgarian may alternate with a possessive or relational adjective in pre-position (*STOLICATA na Italiya Rim* ‘CAPITAL-THE of Italy *Rome*’, *italianskata STOLICA Rim* ‘Italian-the CAPITAL *Rome*’). Alternations of possessive pronouns and possessive pronoun clitics in Bulgarian are also observed. A noun modifier in pre-nominal position can alternate with a prepositional phrase (*ski INSTRUKTORĀT* ‘ski INSTRUCTOR-THE’ vs. *INSTRUKTORĀT po ski* ‘INSTRUCTOR-THE at ski’).

In Greek, the genitive phrase may alternate with the preposition *σε* ‘at’ followed by accusative case, e.g., Γεώργιος Μπαμπινιώτης, Καθηγητής του Πανεπιστημίου Αθηνών ‘*Georgios Babiniotis*, PROFESSOR of the University of Athens’ or Γεώργιος Μπαμπινιώτης, Καθηγητής στο Πανεπιστήμιο Αθηνών ‘*Georgios Babiniotis*, PROFESSOR at the University of Athens’. In this case, the two structures may convey a different meaning. The alternation is not possible for all proper names, e.g., Αλέξης Τσίπρας, Πρωθυπουργός της Ελλάδας ‘*Alexis Tsipras*, PRIME MINISTER of the Greece’, \*Αλέξης Τσίπρας, Πρωθυπουργός στην Ελλάδα ‘*Alexis Tsipras*, PRIME MINISTER in the Greece’. A location proper name describing residency at a continent, a country or a city may alternate as an adjective modifier or a PP complement attached to the trigger of a personal name (the same is true for English, Bulgarian, and Serbian), e.g., Greek: ο Πρωθυπουργός της Ελλάδας ‘the PRIME MINISTER of the Greece’ or ο Έλληνας Πρωθυπουργός ‘the Greek PRIME MINISTER’; Serbian: *AMBASADA Grčke* ‘EMBASSY of Greece’ vs. *Grčka AMBASADA* ‘Greek EMBASSY’ (while in French, the adjective follows the trigger: *le PRÉSIDENT français François Mitterrand* ‘the PRESIDENT of-French.Adj *François Mitterrand*’).

In French, we may have an alternation of the preposition *de* ‘of’ with the preposition *à* ‘at’, e.g., **Martin Vetterli**, *PROFESSEUR de l’École polytechnique fédérale de Lausanne* ‘**Martin Vetterli**, PROFESSOR of the Federal Polytechnic School of Lausanne’ or **Martin Vetterli**, *PROFESSEUR à l’École polytechnique fédérale de Lausanne* ‘**Martin Vetterli**, PROFESSOR at the Federal Polytechnic School of Lausanne’.

In Serbian, syntactic alternations are permissible, to some extent, with organization names: a complement in the genitive case instead of a PP complement (**MINISTARSTVO rada i socijalne politike** ‘MINISTRY of Labor and Social Policy’ instead of a **MINISTARSTVO za rad i socijalnu politiku** ‘MINISTRY for Labor and Social Policy’).

The features of the triggers in the five languages are summarized in Table 10.

## 8 Conclusion

The semantic classification and the syntactic patterns of single and multiword names in Bulgarian, English, French, Greek, and Serbian, may provide reliable data for rule-based Named Entity Recognition (NER).

Linguistic features and distribution facts are used to identify MWEs in NER tasks – both in handcrafted rule-based systems that rely heavily on linguistic knowledge, and in machine-learning techniques. In their research on the application of MWEs and NEs in keyphrase extraction, Nagy T. et al. (2011) conclude that previously known noun compounds are beneficial in NER, and that identified NEs enhance MWE detection, as noun compounds and multiword NEs are linguistically similar and sometimes it is not easy to distinguish between the two.

These arguments are further supported by the tagging practice where both compound nouns and multiword NEs are often tagged as nouns, as their linguistic behaviour is similar to that of single-word nouns (Vincze et al. 2011). Approaches such as that of Nagy T. et al. (2011) also use features involving NEs or pertaining to NEs (i.e., orthography and semantics of keyphrase candidates; positions of a token belonging to a specific NE class, as certain classes of NEs can be identified by their position in the beginning, in the middle or at the end of a keyphrase candidate). Galicia-Haro et al. (2004) discuss the (Spanish) composite NEs (titles of books, movies, songs, etc.) that are described in terms of syntactic and semantic features and of local context and consider discourse features such as introductory words, prepositions, redundancy; specific sets of names, etc.

Rule-based systems usually rely on large-scale lexical resources and grammars, often in the form of regular expressions or Finite State Transducers (Savary & Piskorski 2011; Maurel et al. 2011). Much work has been done on rule-based NER

## 2 Semantic and syntactic patterns of multiword names

Table 10: Comparison of the morphological and syntactic features of the five languages.

Features that concern the trigger	English	Bulgarian	French	Greek	Serbian
adjective in pre-position	+	+	+	+	+
adjective in post-position	-	-	+	-	+
PP in post-position	+	+	+	+	+
genitive phrase in post-position	-	-	-	+	+
genitive phrase in pre-position	-	-	-	+	-
genitive determiner in pre-position	+	-	-	-	-
definite article	+	-	+	+	-
definite morpheme	-	+	-	-	+
obligatory definiteness with modifier Adj/PP in a sentence	+	+	+	+	+
possessive pronoun clitic	-	+	-	+	-
dependencies expressed by prepositions	+	+	+	+	+
dependencies expressed by cases	-	-	-	+	+
analytically expressed dependencies	+	+	+	+	+
genitive determiner and PP alternation	+	-	-	-	-
genitive and PP alternation	-	-	-	+	+
poss. pronoun and possessive clitic alternation	-	+	-	-	-
noun modifier and PP alternation	+	+	-	+	+
Features that concern the whole NE					
pre-position of the trigger	+	+	+	+	+
apposition	+	+	+	+	+
interrogative clitic	-	+	-	-	-

for the five languages discussed in this paper, although machine learning methods prevail. A set of general NER rules with reasonable accuracy has been developed for rule-based annotation of NEs in Bulgarian (Karagiozov et al. 2012), French (Maurel et al. 2011), Greek (Farmakiotou et al. 2000), and Serbian (Krstev et al. 2013). Vitas et al. (2007) discuss semantic and morphological (derivational and inflectional) properties of proper names in Serbian (plus French and English) taking into account the significance of regular derivation and the properties and function of possessive and relational adjectives produced from proper names. Koeva & Dimitrova (2015) discuss a strategy for a linguistic description and classification of Bulgarian NEs referring to persons, and their application in several resources (lexicons and an annotated corpus) for the definition and evaluation of a set of NER rules.

The syntactic patterns presented in this paper are formulated as rules comprising morphological characteristics and syntactic dependencies related to the semantic properties of personal, location and organization NEs in Bulgarian, English, French, Greek, and Serbian. We intend to further exploit the formally encoded linguistic information in rule-based NER approaches. Moreover, as the syntactic patterns for different languages are linked to the same semantic pattern, they can be considered equivalent at the conceptual level and may be applied to any task that involves multilingual processing: cross-lingual information extraction and text classification, multilingual summarization and machine translation. Last but not least, the presented approach contributes to comparative language studies and may be further extended to other word classes that show relatively regular morphological properties and syntactic dependencies.

## Abbreviations

ADJ	adjective	GENP	(noun) phrase in genitive
ADV	adverb		
AFF	affiliation	INDEFTRIGGER	indefinite trigger
DEFART	definite article	INS	instrumental
DEFADJ	definite adjective	LIT.	literal translation
DEFPOSSPRON	definite possessive pronoun	LOC	location
		MWE	multiword expression
DEFTRIGGER	definite trigger		
(DOM)SPEC	(domain) specification	MWLOCN	multiword location name
GENDET	genitive determiner	MWORGN	multiword organization name

NE	named entity	POSSADJ	possessive adjective
QUCL	interrogative clitic	(POSS)CL	(possessive) pronoun clitic
NER	named entity recognition	POSSPRON	possessive pronoun
PERN	personal name	PP	prepositional phrase

## References

- ACE. 2018. *Automatic content extraction. English entity guidelines v6.6 15 2008.06.13*. ACE. <https://www ldc.upenn.edu/sites/www ldc.upenn.edu/files/english-entities-guidelines-v6.6.pdf>, accessed 2018-4-19.
- Anderson, John. 2007. *On the grammar of names*. Oxford University Press.
- Carroll, John. 1985. *What's in a name? An essay in the psychology of reference*. WH Freeman/Times Books/Henry Holt & Co.
- Farmakiotou, Dimitra, Vangelis Karkaletsis, John Koutsias, George Sigletos, Constantine D. Spyropoulos & Panagiotis Stamatopoulos. 2000. Rule-based named entity recognition for Greek financial texts. In *Proceedings of the Workshop on Computational Lexicography and Multimedia Dictionaries (COMLEX 2000)*, 75–78.
- Fleischman, Michael & Eduard Hovy. 2002. Fine grained classification of named entities. In *Proceedings of the 19th International Conference on Computational Linguistics*, vol. 1, 1–7.
- Galicia-Haro, Sofia, Alexander Gelbukh & Igor Bolshakov. 2004. Recognition of named entities in Spanish texts. *MICAI 2004: Advances in Artificial Intelligence*. 420–429.
- Giuliano, Claudio. 2009. Fine-grained classification of named entities exploiting latent semantic kernels. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, 201–209.
- Huddleston, Rodney. 1988. *English grammar: An outline*. Cambridge University Press.
- Karagiozov, Diman, Anelia Belogay, Dan Cristea, Svetla Koeva, Maciej Ogródniczuk, Polivos Raxis, Emil Stoyanov & Cristina Vertan. 2012. I-Librarian—Free online library for European citizens. *INFOtheca* 13(1). 27–43.
- Koeva, Svetla & Tsvetana Dimitrova. 2015. Rule-based person named entity recognition for Bulgarian. In *Slavic languages in the perspective of formal grammar (Proceedings of FDSL 10.5, Brno 2014)*, vol. 37 (Series Linguistik International), 121–139. Peter Lang.

- Krstev, Cvetana, Ivan Obradović, Miloš Utvić & Duško Vitas. 2013. A system for named entity recognition based on local grammars. *Journal of Logic and Computation* 24(2). 473–489.
- Magnini, Bernardo, Matteo Negri, Roberto Prevete & Hristo Tanev. 2002. A WordNet-based approach to named entities recognition. In *Proceedings of the 2002 Workshop on Building and Using Semantic Networks*, vol. 11, 1–7. Stoudsburg: Association for Computational Linguistics.
- Maurel, Denis, Nathalie Friburger, Jean-Yves Antoine, Iris Eshkol & Damien Nouvel. 2011. Cascades de transducteurs autour de la reconnaissance des entités nommées. *Traitement automatique des langues* 52(1). 69–96.
- McDonald, David. 1996. Internal and external evidence in the identification and semantic categorization of proper names. *Corpus processing for lexical acquisition*. 21–39.
- Nagy T., Istvan, Gábor Berend & Veronika Vincze. 2011. Noun compound and named entity recognition and their usability in keyphrase extraction. In *RANLP*, 162–169.
- Rodríguez, Horacio, Salvador Climent, Piek Vossen, Laura Bloksma, Wim Peters, Antonietta Alonge, Francesca Bertagna & Adriana Roventini. 1998. The top-down strategy for building EuroWordNet: Vocabulary coverage, base concepts and top ontology. In *EuroWordNet: A multilingual database with lexical semantic networks*, 45–80. Springer.
- Savary, Agata & Jakub Piskorski. 2011. Language resources for named entity annotation in the National Corpus of Polish. *Control and Cybernetics* 40. 361–391.
- Searle, John R. 1958. Proper names. *Mind* 67(266). 16–173.
- Sekine, Satoshi & Chikashi Nobata. 2004. Definition, dictionaries and tagger for Extended Named Entity Hierarchy. In *4th International Conference on Language Resources and Evaluation (LREC)*, 1977–1980. Lisbon, Portugal.
- Vincze, Veronika, Istvan Nagy T. & Gábor Berend. 2011. Multiword expressions and named entities in the Wiki50 Corpus. In *Proceedings of Recent Advances in Natural Language Processing*, 289–295. Hissar, Bulgaria.
- Vitas, Duško, Cvetana Krstev & Denis Maurel. 2007. A note on the semantic and morphological properties of proper names in the Prolex project. *Lingvisticae Investigationes* 30(1). 115–133.
- Zhang, Jie, Dan Shen, Guodong Zhou, Jian Su & Chew-Lim Tan. 2004. Enhancing HMM-based biomedical named entity recognition by studying special phenomena. *Journal of biomedical informatics* 37(6). 411–422.