IDENTIFYING THE HEAD IN TOPONYMIC CONSTRUCTIONS

DENIS PAPERNO

0. THE NOTION OF A TOPONYMIC CONSTRUCTION. In this paper, I will call two nouns or noun phrases a *toponymic construction* if the following is true:

- one of them (A) is a toponym;
- the other one (B) specifies the class of geographical objects to which A belongs;
- A and B occur within the same noun phrase;
- the fact that A is the name for B is not stated in the noun phrase that contains them.

We will refer to the element that specifies the class of geographical object (B) as the *generic component*.

Accordingly, *the city of Moscow*, *New York State* and *Lake Michigan* are toponymic constructions (although it may be argued that the last two are proper names). The following expressions are not: *Moscow; New York, city, the state, beautiful country, this wonderful Moscow, New York is big* (lack of one of the components); *Everest is a mountain, this city is called Kiev* (the components are not in the same NP); and *the village named Ouagadougou* (the naming relation is stated explicitly).

We also leave aside phrases with parenthetical specifications like *New York, the city that never sleeps*, showing properties significantly different from those of usual toponymic constructions, e.g. free order of components (*the city that never never sleeps, New York*). Of course, the semantic difference between a parenthetical specification and a normal component of a toponymic construction is very subtle, but it still exists.

In many of the world’s languages, the syntactic relationship between the components of a toponymic construction is of the same type as in some of the language’s possessive constructions. Toponymic constructions that show the same syntactic properties as possessive phrases may be called *genitive toponymic constructions*, as opposed to *appositive toponymic constructions*.

A genitive toponymic construction can be marked in a number of ways, e. g. with a preposition (Spanish ciudad de Madrid ‘the city of Madrid’), an izafet marker (Persian shahr-e Tehran ‘Tehran city’), a genitive case marker (Finnish Helsingi-n kaupunki ‘the city of Helsinki’), a possessivity marker (Hungarian Moszkva varos-a ‘Moscow city’), a possessive adjective suffix (Old Church Slavonic гора елеоньска ‘Mount Elaeon’) or simply null (Komi Ухта кар ‘Ukhta city’, which shows the same syntactic properties as Вася керка ‘Vasya’s house’).

1. IDENTIFYING THE HEAD OF THE PHRASE. It appears that within genitive toponymic constructions, the generic component is always the head of the noun phrase, just as it is in the corresponding possessive construction, e.g. *the city of Moscow* and *the father of Mary*. For appositive constructions this issue is not so clear.

To deal with this, we make the following assumptions: 1) one of the components (the generic or proper term) is the head of the phrase; 2) the construction is endocentric (that is, the dependent component can be omitted without affecting the grammaticality of the sentence).

The omission test resolves the issue in most cases and identifies the generic component as the head of the phrase. This component resists being omitted for a number of reasons:

1. the generic component may control the agreement of the noun phrase as a whole (Serbo-Croatian Grad [M] Sarajevo [N] je lijep [M] / Sarajevo je lijepo [N]
   ‘Sarajevo city/ Sarajevo is beautiful.’);
2. the generic component, unlike the toponym, may co-occur with the article (Afrikaans Ek woon in die dorp Stellenbosch / in Stellenbosch
   ‘I live in Stellenbosch village.’ / ‘in Stellenbosch.’);
only the generic component, not the toponym, takes the article or case affixes required by the noun phrase (Albanian\(^1\) \textit{qytet-i [river-DEF] Elbasan} ‘Elbasan river’ / \textit{Elbasan-i [Elbasan-DEF]}; Moksha Mordva \textit{мон качень касонь Orydu веле-са [LOC] / Orydu-са [LOC]} ‘I was born and grew up in Achadovo village.’ / ‘in Achadovo.’) 

I am aware of only three cases where the head is the toponym. The first example is an English toponymic construction where the generic component is preceded by the toponym. Other than in established proper names (like New York City), the distribution of this construction is restricted to the names of states (regarded ungrammatical in isolation, they are only accepted by the native speakers in order to disambiguate the context, e.g. “Did you see the White House when you were in Washington?” – “I was in Washington state, not the city.”\(^{Kansas state}\), though sounding strange for some speakers, is also attested, but not \(*\textit{California state}\), cities (e.g. \textit{San Francisco city}, which is attested, though some speakers find it strange) and some others (e.g. \textit{Hudson river}, where the omission of the toponym leads to ungrammaticality unless the article is inserted: \textit{It’s unusual to see elephants in Washington city/ in the city/ \*in the Washington city}. Note that with the other toponymic constructions of English, the article is necessary: \textit{Hudson river, Boston city, but the river Hudson, the city of Boston} (here the generic component is the head of the phrase).

Another example is taken from Adygh\(^2\), where the generic component takes a special case marking: \textit{(Псыхъо-у [river-ADV]) Москва мьчыкъэу цыы ‘Moscow (river) is not far.’ / Псыхъо мьчыкъэу цыы ‘River is not far.’}

The third example is provided by modern Chinese, where special lexemes (not allowed elsewhere) are used within a toponymic construction: \textit{chengshi ‘city’ / Taipei (*cheng)shi ‘Taipei city’}.

Our assumptions are not sufficient to identify the head of appositive toponymic constructions in Spanish and Portuguese. In these languages substantives don’t have morphologically marked case; both toponyms and generic components require an article (as all common nouns do); the gender of the components of the appositive construction is always the same (Spanish \textit{estoi en el [M] lago [M] Titicaca [M] / en el Titicaca / en el lago} ‘I’m at Titicaca lake.’ / ‘at Titicaca.’ / ‘at the lake.’) Examples from these languages seem structurally ambiguous.

2. THE ORDER OF THE COMPONENTS. Let’s set aside for the moment the non-productive constructions that may be regarded as proper names, attributed to specific toponyms, e.g. the Russian \textit{Москва-река} and the English \textit{Lake Michigan}. An interesting observation can be made regarding toponymic constructions, both genitive and appositive. In left-branching (“postpositional”) languages the head’s unmarked\(^3\) position is to the right of the dependent component, while in right-branching (“prepositional”) languages it is to its left. Spanish and Portuguese, that are right-branching languages and put toponyms to the right of the generic component, fit this generalization if we assume the generic component to be the head (this is obvious for the genitive constructions, but the appositive ones are ambiguous). The generalization holds, no matter which of the components (the toponym or the generic component) is the head.

Toponymic constructions are similar to pre- and postpositional phrases and possessive phrases in one other respect: the components of toponymic constructions can be demarcated by the same set of elements that the corresponding possessive constructions are demarcated by: \textit{the city of Boston, like the mother of Mary}; in Russian, the poetically marked город \textit{прекрасный}.

---

1 Plank 2003: 34.

2 The Adygh data were kindly provided by Yu. Lander.

3 Some languages allow the reverse order as well (Avar \textit{Махачкъала шагъар/шагъар Махачкъала} ‘the city of Makhachkala’); such languages are not numerous.
Москва ‘the beautiful city of Moscow’ like город прекрасный Петра ‘the beautiful city of Peter.’

In other words, it seems very natural to regard the dependent component of the toponymic construction as filling a complement position of the head. One can assume that the generic component has a semantic valence for a toponym, realized by this position. Correspondingly, one could assume that in English, Chinese, and Adyghe toponyms have a semantic valence for a generic specifier.

The question why the structure for toponymic expressions in a given language is what it is, with the given proportion of genitive and appositive, toponym-headed and generic-headed constructions is a separate issue that I leave without analysis for now.

REFERENCE