

Proper names: a translation corpus view

Bert Le Bruyn^a, Olga Borik^b, Ljudmila Geist^c, Jianan Liu^a, Sadhwi Srinivas^d, Hagay Schurr^e and Daria Seres^f

Universiteit Utrecht^a, Universidad Nacional de Educación a Distancia^b, Universität Stuttgart^c, William & Mary^d, The City University of New York^e, Universität Graz^f

To contribute to the cross-linguistic investigation of the syntax-semantics interface of proper names (PNs) and of reference more generally, we are conducting a translation corpus study focusing on the PNs in the first three chapters of *Harry Potter and the Philosopher's Stone* (HPPS) and their translations to a variety of languages. We present §1) the different construction types (CTs) with proper names in the English original, §2) our selection of target languages, and §3) a preliminary analysis of the cross-linguistic variation we find in two of the CTs: *title + name* (e.g., *Professor McGonagall*) and *kinship term + name* (e.g., *Uncle Vernon*), both in their regular argument use. §1 and §2 establish the scope of our study. §3 argues for its theoretical relevance.

1. English: CTs with proper names | The first three chapters of HPPS contain 701 occurrences of PNs. For our English data exploration, we analyzed these and subdivided them into 60 descriptive CTs, covering form and use. (1) to (10) present a sample of regular argument CTs:

(1) 'a' + modifier + proper name	<i>A screaming Dudley</i>	(6) place name with definite article	<i>The Isle of Wight</i>
(2) 'that' + modifier + proper name	<i>That odd Harry Potter</i>	(7) Mrs. + name	<i>Mrs. Dursley</i>
(3) 'her' + affectionate name variation	<i>Her Ickle Dudleykins</i>	(8) Mrs. + adjective + noun	<i>Mrs. Next Door</i>
(4) 'the' + family name _{plural}	<i>The Dursleys</i>	(9) title + name	<i>Professor McGonagall</i>
(5) 'the' + kind name _{plural}	<i>The Muggles</i>	(10) kinship term + name	<i>Uncle Vernon</i>

Even though bare PNs in regular argument positions cover a good deal of our data (n=339), most PNs in our corpus turn out to be either modified (see (1) to (10)) or not to occur as regular arguments. Next to the vocative and naming uses, we also find PNs as modifiers (e.g., *Harry Potter Day*), incorporated objects (e.g., *Harry-hunting*), place names in addresses (e.g., *Cokeworth*) and preceded/followed by bound morphemes (e.g., *unDursleyish*). With this wide variety of CTs, our study goes well beyond the empirical scope of the current literature.

2. Languages in our study | For our cross-linguistic data exploration, we selected 100 occurrences of PNs in English, covering all CTs. Data collection is ongoing, but we have now covered 9 languages. Our selection includes 'article-less' languages (Mandarin, Russian), languages with a definite but without an indefinite article (Hebrew, Macedonian), and languages with definite and indefinite articles, some with a generalized use of definite articles with PNs (Catalan, Greek), others without (German, Norwegian, Italian). Our data replicate the basic co-occurrence generalizations of definite articles and PNs that we know from the literature:

- (11) Dudley > Dá lì (Mandarin), Dadli (Russian), Dadli (Hebrew), Dadli (Macedonian), Dudley (German), Dudleif (Norwegian), Dudley (Italian), el Dudley (Catalan), o Ntántli (Greek)

Another generalization that is replicated in our data is that the definite article that co-occurs with PNs in Catalan and Greek disappears in naming constructions (Matushansky 2009):

- (12) fills que es deien _Harry (lit. 'sons that were named Harry') (Catalan)

- (13) o anipsiós tou legótan _Chári (lit. 'the nephew his was _called Harry') (Greek)

We argue that our selection of languages is useful to probe the nature of PNs across languages and that of reference in general. The appearance of definite articles has been a core argument in favor of quotation theories of PNs. By collecting rich data (§1) from languages with varied uses of definite articles (§2), we cannot but shed new light on the nature of PNs and that of reference. As a proof of concept, we analyze the translations of the CTs in (9) and (10).

3. The title and kinship constructions | We refer to (9) as the *title construction* (TC, De Belder 2009, 2022) and to (10) as the *kinship construction* (KC). (14) and (15) summarize our data (see §4 for Hebrew and Mandarin). The TC exists in two variants, both with the order title+PN, the

difference residing in the absence (a)/presence (b) of a definite article in front of the title noun (suffixed in Macedonian). The KC exists in the same two variants. We find inter-linguistic variation: Russian, German and Norwegian TCs and KCs appear without definite articles vs. Catalan and Greek TCs and KCs appear with definite articles. We also find intra-linguistic variation: Macedonian and Italian have TCs with definite articles but KCs without.

(14) *Title construction (TC)*

(a) Professor McGonagall Russian, German, Norwegian

(b) The professor McGonagall **Macedonian, Italian**, Catalan, Greek

(15) *Kinship construction (KC)*

(a) Uncle Vernon Russian, German, Norwegian, **Macedonian, Italian**

(b) The uncle Vernon Catalan, Greek

Our data raise two questions: (i) why is the order ‘common noun + PN’? (ii) how to analyze the inter-/intra-linguistic variation? Question (i) directly touches on the nature of PNs. If we take PNs – in line with quotation theories – to be predicative, and if we make the plausible assumption that common nouns are also predicative, there does not seem to be a principled reason to expect the specific order we find in the TC and KC. Indeed, if title/kinship terms and PNs are both predicates, the simplest way to compose their meanings would be to rely on predicate modification, an operation that does not require any specific order. To remedy, we generalize De Belder’s analysis of the TC and assume that the TC and KC start life as argument-predicate structures with the PN as the (subject) argument and the common noun as the predicate. Further following De Belder, we assume the common noun raises, leading to the observed surface orders.

For De Belder, the common noun by default raises all the way to D, in line with Longobardi’s original ideas about N-to-D raising (Longobardi 1994). We propose (16) as the semantic correlate of raising in De Belder’s analysis. With *Professor* as input, the output combines with *McGonagall*, and refers to the unique individual who is a professor and identical to *McGonagall*.

(16) $P_{\langle e,t \rangle} \Rightarrow \lambda y_e \lambda x_e (P(x) \& x=y)$

Even though De Belder’s analysis works well for the (a) variants in (14) and (15), it fails to account for inter- and intra-linguistic variation. This brings us to Question (ii). To account for inter-linguistic variation, several semantic and syntactic solutions are possible, but for the intra-linguistic variation we find between the TC and the KC in Macedonian and Italian, a more innovative take on the syntax-semantics interface of reference is needed. Inspired by recent work on speech acts at the propositional level (Wiltschko 2021; Krifka 2023), we hypothesize that speech act participants have a dedicated projection in an extended DP-layer to which we refer as SPP. SPP constitutes a relevant landing site for relational kinship nouns like *uncle* but not for sortal nouns like *Professor*. We propose that the contrast between the TC and KC in Macedonian and Italian resides in the relational/sortal distinction of their common nouns: relational kinship nouns raise to SPP, saturate their relational argument and undergo the operation in (16). Sortal nouns in these languages can raise (e.g., to a NUMBERP or a GENDERP) but stay below the extended DP-layer and can only undergo a variant of (16) that does not lead to an argumental type after combining with a PN as in (17), hence requiring the addition of a definite article.

(17) $P_{\langle e,t \rangle} \Rightarrow \lambda y_e \lambda x_e (P(x) \& x=y)$

Semantically speaking, support for a contrast between the TC and KC comes from the fact that the latter but not the former requires a link with a speech act participant (SPA), be it a direct SPA or the character whose perspective the narrator chooses to tell the story.

4. Conclusion and outlook | We have presented the empirical scope of our translation corpus study (§1,§2) and we have provided a proof of concept to support its theoretical relevance (§3).

We note that the analysis we presented of the TC and KC is preliminary. Empirically, we only covered a small sample of our data (Mandarin relies on the ‘PN+common noun’ order in the TC and KC, the Hebrew data suggest an additional sensitivity to verb placement, and the superficially similar construction in (7) leads to yet a different cross-linguistic interaction with definite articles). Methodologically, our approach requires replication and triangulation.