The theory of argument formation: between kinds and properties *

Jianan Liu
Utrecht University

Shravani Patil
Tilburg University

Hagay Schurr
The Graduate Center, CUNY

Daria Seres
University of Graz

Olga Borik
Universidad Nacional de Educación a Distancia

Bert Le Bruyn
Utrecht University

Abstract Chierchia (1998) developed a cross-linguistic extension to Carlson's seminal work on bare nouns (BNs), producing the most influential theory of argument formation to date, henceforth the Kinds Approach (KA). The core achievements of the KA included the derivation of the generalized narrow scope behavior of BNs and of the existence of generalized classifier languages. There are cracks in the picture, though. The narrow scope behavior of BNs is more fine-grained than is generally assumed and the KA lacks the flexibility to deal with it (Le Bruyn & Swart 2022). The appeal of the KA's derivation of the existence of generalized classifier languages heavily relied on all nouns in these languages being mass-like, an assumption that has since been abandoned (Chierchia 2010; Jiang 2020). These developments call for a reassessment of the KA and one of its closest competitors: Krifka (2003). Krifka assumes nouns never start life as kinds but as predicates, leading us to qualify his approach as a Properties Approach (PA). We adopt a translation corpus methodology and assess the explanatory potential of the KA and the PA by comparing the distribution of BNs and related expressions in (in)definite contexts across six typologically different languages. Our results show that the PA has a distinct advantage over the KA and identify pseudo-incorporation and the way it varies across languages as a primary focus for future research.

Keywords: bare nouns, kind reference, (in)definiteness, parallel corpora

1   Introduction

Chierchia's (1998) cross-linguistic extension to Carlson's (1977) kinds analysis of English bare plurals (BPs) (henceforth the *Kinds Approach* (KA)) is the most influential theory of argument formation to date. Among the core facts that it was able to derive, we mention the generalized narrow scope behavior of bare nouns (BNs) and the existence of generalized classifier languages. After Chierchia 1998, the KA was further developed and a number of competing theories were proposed. None of them, however, have achieved anything close to the same popularity.

We single out Krifka 2003 as the KA's conceptually closest competitor. Both the KA and Krifka's approach are cast in a type-shifting framework along the lines set out in Partee 1987. However, while Krifka's approach takes nouns to uniformly start life as predicates (hence we qualify it as a *Properties Approach* (PA)), the KA does not extend this to all languages, as we discuss in Section 2.1. The other crucial difference between the two lies in the fact that predicates-to-arguments shifts are not ranked in the PA. The PA consequently does not attribute any special status to predicates-to-kinds shifts, differently from the KA.

Two developments related to the core facts mentioned above invite an open-minded reassessment of the explanatory potential of the KA and the PA.

On narrow scope

Even though the narrow scope behavior of BNs is often considered to require an approach with a central role for kinds, Krifka (2003) was one of the first to argue that all one needs is a locality requirement on type-shifting, an assumption that is also central to the KA. The exact implementation of the locality requirement is slightly different in the two approaches, though. Comparing the narrow scope accounts of both approaches, Le Bruyn & Swart (2022) find that they differ in their predictions about the scope of scrambled BNs. Whereas the KA predicts no difference between scrambled and unscrambled BNs, the PA predicts the former to be able to take wide scope. Le Bruyn & Swart (2022) argue that scrambled bare plurals in Dutch take wide scope over negation and consider this to be an empirical argument in favor of the PA.

On classifier languages

One of the most appealing achievements of Chierchia's original version of the KA was that the existence of generalized classifier languages like Mandarin followed directly from the notion that some languages can have their nouns uniformly start life as kinds. Krifka was able to mimic the need for classifiers in Mandarin but needed to do so in the lexical entries of individual nouns, a *prima facie* less attractive move. However, after developmental psychologists and linguists had criticized the original version of the KA for ignoring the distinction in Mandarin between

mass and count nouns, Chierchia (2010) was led to refine his theory and assume that nouns are lexically marked to start life as mass or count kinds (see also discussion in Jiang 2020). The crucial point here is that the KA does not escape the need to adopt the same type of lexicalist approach as the PA, arguably leveling the playing field between the two.

Since these developments suggest that two major achievements of the KA have not fully withstood the test of time, we find that the PA comes out as a relevant competitor, which calls for an open-minded reassessment of the predictions of both approaches.

Methodologically, we want to cast the net wide, and we do so in two respects. First, we consider a sample of six languages. In addition to four that represent the same languages or language families that appeared in Chierchia (1998), *viz.* Spanish (Romance), German (Germanic), Russian (Slavic), and Mandarin (Sino-Tibetan), we added Hindi for its pivotal role in the KA (for an exploration of bare nominals in Hindi, an 'article-less' language, and its possible implications for our understanding of kind reference, see Dayal 2004), and Hebrew as a distinct in-between language type with its definite but no indefinite article (Doron 2003), which sets it apart from both 'article-less'and article languages in the rest of our sample. Second, we do not focus on preset examples but rather rely on what the analysis of a (small) translation corpus brings us, *viz.* the translations of the first chapter of J.K. Rowling's *Harry Potter and the Philosopher's Stone* (henceforth *HP*). Translation corpus research has recently gained traction as a valuable tool in the cross-linguistic semanticist's toolbox next to questionnaires and experimental methodologies (see, e.g., Bremmers, Liu, van der Klis & Le Bruyn 2022; Mulder, Schoenmakers, Hoenselaar & de Hoop 2022; Gehrke 2022; Klis, Le Bruyn & Swart 2022).

The increasing number of papers that make use of translation corpus methodology in cross-linguistic semantics attests to the gradual but steady maturation of this subfield (Le Bruyn, Fuchs, van der Klis, Liu, Mo, Tellings & Swart 2022; Le Bruyn, Kliss & Swart to appear; Le Bruyn & de Swart submitted). In addition to the maximal semantic comparability of parallel translations, we find the translation corpus approach on the basis of HP particularly attractive for two additional reasons: (i) the English original gives us a reasonable grip on the interpretation of BNs in the translations, even for languages that do not overtly mark (in)definiteness; (ii) the corpus can easily be extended, both in number of languages (HP was translated to over 80 languages) and in number of words (HP is a 7-volume series that makes for a corpus of approximately one million words, if used in its entirety, and we are only looking into the first chapter of the first volume).

The paper is organized as follows. In Sections 2 and 3, we work out the predictions the KA and PA make for the availability of BNs in argument position in

singular/plural (in)definite contexts. In Section 4, we present the predictions of the KA and PA to be tested against the results of our parallel corpus study. We also add a brief note on pseudo-incorporation. Section 5 concludes the paper.

2 The Kinds Approach and its predictions

The KA draws on Carlson's (1977) influential work on BPs in English for the basic intuition that kind and indefinite readings of BNs are related, a common thread throughout the history of the KA. Next, we work out the hypotheses that make up the KA. Given that the approach typically distinguishes between classifier and non-classifier languages, we discuss these in turn in Sections 2.1 and 2.2. In Section 2.3, we work out the predictions the KA makes for the availability of BNs in argument position in (in)definite singular/plural contexts in the languages of our sample.

2.1 Non-classifier languages

For non-classifier languages, we first sketch the approach proposed in Chierchia 1998 and then highlight the refinements that have been implemented since.

2.1.1 Chierchia (1998)

Chierchia (1998) proposes that non-classifier languages come in two guises, related to a parameter he calls the *Nominal Mapping Parameter*:

[-arg,+pred]: nouns uniformly start life as properties and need to rely on overt or covert functional material to shift to argumental types;

[+arg,+pred]: nouns start life as properties or kinds and can shift between argumental and non-argumental types without the intervention of overt or covert functional material.

Building on Partee 1987, Chierchia assumes a type-shifting framework in which nouns can flexibly shift between types. The basic type-shifts he assumes are: the *iota* shift (ι, from type $<e,t>$ to type e), the existential shift ($\exists$, from type $<e,t>$ to type $<<e,t>,t>$), the *down* shift ($\cap$, from type $<e,t>$ to kinds, type e_k), and the *up* shift ($\cup$, from kinds to their instantiations). Chierchia hypothesizes that shifts to argumental types are constrained by the type-shift ranking in (1) and the Blocking Principle in (2), the former prioritizing the *down* shift over the *iota* and the existential shifts, the latter proscribing shifts from applying covertly in languages that have lexicalized the shifts in their determiner systems. Definite articles are typically thought to lexicalize the *iota* shift and indefinite articles the existential shift.

(1)   Type-shift ranking in the KA (to be adapted): $\cap > \{\iota, \exists\}$

(2)   *The Blocking Principle*
       For any type shifting operation τ and any $X : {}^*\tau(X)$ if there is a determiner
       D such that for any set X in its domain, $D(X) = \tau(X)$

Outside standard type-shifting, Chierchia assumes kinds can give rise to derived indefinite readings through *Derived Kind Predication* (DKP), as defined in (3):

(3)   *Derived Kind Predication* (DKP)
       If P applies to objects and k denotes a kind, then $P(k) = \exists x \, [{}^{\cup}k(x) \& P(x)]$

DKP is Chierchia's formalization of the link between kind and indefinite readings of BNs and is the key to deriving the generalized narrow scope behavior of BNs (see Section 1). In essence, it take a kind k and the predicate P that the k combines with, and returns the proposition that asserts the existence of individuals that are members of the kind and satisfy the descriptive content of the predicate. DKP is assumed to apply locally, making sure that the existential quantifier it generates always takes the narrowest possible scope.

2.1.2   After Chierchia (1998)

Where Chierchia (1998) distinguishes two types of non-classifier languages, we find a reduction to a single type in the later literature, *viz.* [-arg,+pred] (Dayal 2004; Chierchia 2010; Jiang 2020). The original motivation for the distinction between two types came from the contrast between languages like Italian and English: the former restricts the syntactic positions in which its BPs can appear whereas the latter does not (see also Longobardi 1994). However, the current consensus attributes this fact to syntax, essentially rendering redundant any attempt to capture it with a semantic parameter. For example, Jiang (2020) derives the Italian/English opposition from the syntactic parameter $\pm \, \text{ARG}_{\text{unrestricted}}$.

A further evolution lies in the ranking of type-shifts. Chierchia originally proposed that the *down* shift should be ranked above both the *iota* and the existential shift. Starting from Dayal 2004, the consensus seems to be that the *iota* shift should be unranked with respect to the *down* shift and that both should be ranked above the existential shift (see also Jiang 2020), resulting in the final ranking in (4):

(4)   Type-shift ranking in the KA (final): $\{\cap, \iota\} > \exists$

For languages with definite articles, this change does not affect the predictions made by the KA, as the Blocking Principle (2) independently blocks the *iota* shift for BNs. For languages like Hindi, with no definite article, this change entails that

nouns may directly undergo both the *down* shift – leading to a kind reading – and the *iota* shift – leading to a definite reading.

Lastly, researchers within the KA have increasingly worked out how number interacts with kinds. Chierchia (1998) already hypothesized that the *down* shift requires plural nouns. Dayal (2004) emphasizes that the same holds for the *up* shift ($^{\cup}$) in number-marking languages and restricts it to kinds built from *plural* nouns. The *up* shift is a crucial ingredient of DKP. With the restriction of the *up* shift to plural kinds, DKP can give rise to indefinite readings for kinds built from plural nouns but not for kinds built from singular nouns.

2.2 Classifier languages: the case of Mandarin

The KA assumes that classifier languages are [+arg,-pred] languages and that their nouns start life as kinds. As we indicated in Section 1, the KA has been refined in that it now recognizes both count and mass kinds. This refinement has no direct impact on the predictions the KA makes about the availability of BNs in argument position, though.

For Mandarin, the availability of BNs in argument position in definite and indefinite contexts is worked out in most detail in Jiang 2020. Definite readings of BNs are derived through *Situation Restriction* (SR), as defined in (5).[1] It takes a kind and returns the maximal member instantiating it in a situation *s*. Indefinite readings are derived through Derived Kind Predication (DKP), as defined in (3).

(5)   *Situation Restriction* (SR):
      $[N_{<e_k>}]_s \rightarrow [N_{<e>}]$ =the maximal member instantiating $N_{<e_k>}$ in a situation *s*

We note that Mandarin is a classifier language that does not mark number. Number considerations are consequently not assumed to play a role in the availability of DKP in this language.

2.3 Predictions

With the KA's hypotheses in place, we can work out the predictions this approach makes for the availability of BNs in argument position in (in)definite singular/plural contexts for the languages under investigation. We will do so in two steps, first working out the predictions for non-classifier and classifier languages in general, and then fine-tuning them for the languages in our sample based on the articles the languages are assumed to have.

The predictions we will arrive at disregard the possible occurrence of BNs in pseudo-incorporation constructions (we return to this in Section 4). In this section

---

1 The reader is referred to Jiang 2020 for further details.

and in Section 3, we focus on standard argumental BNs, as we assume that they can be distinguished from pseudo-incorporated BNs in a principled way.

2.3.1   General predictions

Assuming together with the KA's proponents that BNs in non-classifier languages start life as type <e,t> expressions and that they can be singular or plural, BNs in definite contexts are predicted to be available through the *iota* shift except for languages that have definite articles. This prediction follows from the type-shift ranking in (4), whereby the *iota* shift is not outranked by other type-shifts, and from the Blocking Principle in (2) – the hypothesis that overt type-shifts (e.g., definite articles) block their covert application. For indefinite contexts, the predictions differ for singular and plural BNs. Indefinite readings of singular BNs are excluded, independently of indefinite article morphology: due to the outranking of the existential type shift by the *iota* and *down* shifts (see (4)), and the number restrictions on the latter two, neither the existential shift nor DKP can derive indefinite readings of singular BNs (see (3)). As far as plural BNs are concerned, the existential type-shift is unavailable, but the *down* and *up* shifts can derive indefinite readings in tandem if the *down* shift is followed by DKP. Our understanding of DKP within the KA is that it is not a type-shift; consequently, it is insensitive to the Blocking Principle. We therefore expect no formal semantic restriction on the availability of BNs in plural indefinite contexts (but see Section 2.2 for a potential contribution from syntactic factors in the form of Jiang's (2020) $\pm$ARG$_{\text{unrestricted}}$ parameter).

For classifier languages, the KA yields the hypothesis that BNs start life as kinds (type e_k). BNs are predicted to be available in indefinite and definite contexts. Definite readings are derived through Situation Restriction (see (5)) and indefinite readings are derived through DKP. As before, we assume that DKP is not a regular type-shift and is insensitive to the Blocking Principle. All other things being equal, the prediction the KA makes for indefinite readings of BNs in classifier languages is thus that they are always available. Whether or not SR interacts with the Blocking Principle (or an adapted version thereof), is an open question we address in Section 4.4. For now, we assume that SR is not sensitive to the Blocking Principle and is consequently always available according to the KA.

2.3.2   Language-specific predictions

To finetune the predictions for the languages in our sample, we must spell out the articles that the KA assumes for the different languages. We do so in (6):

(6)   Assumptions about articles in the KA
        No article: Mandarin, Hindi and Russian (Chierchia 1998; Dayal 2004;

Jiang 2020).
Definite article: Hebrew (Doron 2003).
Definite article$_{SG/PL}$ and indefinite article$_{SG}$: Spanish and German (Chierchia 1998; Dayal 2004; Jiang 2020).

The list in (6) is more extensive than it would need to be since the existence of definite articles has no impact on the predictions of the KA for classifier languages based on the general predictions we worked out above. The same holds for indefinite articles in both classifier and non-classifier languages. We still opt for an exhaustive list because we will assume the descriptive adequacy of (6) to finetune the general predictions of the PA (see Section 3), which are sensitive to a broader range of articles. Relying on a single set of assumptions about articles seems to be the best way to come to a first balanced assessment of the KA and the PA. In Section 4.4, we will explore the impact that slight modifications of these assumptions have on the explanatory potential of the two approaches.

With the assumptions in (6) in place, we can finetune the general predictions made above. There are no changes for classifier languages nor for the availability of BNs in indefinite contexts in non-classifier languages. As before, Mandarin is predicted to allow for BNs in definite and indefinite contexts alike and in non-classifier languages, BNs are expected to be acceptable in plural indefinite contexts, but excluded in singular indefinite contexts. The crucial refinements concern the definite domain in non-classifier languages where the assumptions in (6) lead the KA to predict that Hindi and Russian allow for BNs whereas Hebrew, Spanish and German do not. These language-specific predictions are summarized in Table 1.

We mark the expected availability of BNs per context for each language in Table 1 using the ✓ sign against a green background (=BNs expected to be available) and the ✗ sign against a red background (=BNs expected to be unavailable).

3   The Properties Approach and its predictions

As we indicated in Section 1, the PA is cast in the same type-shifting framework as the KA, but differs from it in assuming that nouns always start life as predicates and that type-shifts are unranked with respect to one another. The PA furthermore does not allow for DKP, arguing that it involves a *sequence* of type-shifts – the *up* and the existential shifts – and consequently violates the consensus that type-shifts are last resort operations. The PA does take over the Blocking Principle as a way to restrict covert type-shifting.

The above hypotheses lead the PA to predict that every language allows its BNs to undergo the *down*, *iota* and existential shifts unless they have articles to block these type-shifts from applying covertly. The upshot is that the availability of definite and indefinite readings of BNs is fully determined by the availability of articles.

	Singular INDEF.	Singular DEFINITE	Plural INDEF.	Plural DEFINITE
Non-classifier languages				
Spanish BSs and BPs	✗	✗	✓	✗
German BSs and BPs	✗	✗	✓	✗
Russian BSs and BPs	✗	✓	✓	✓
Hindi BSs and BPs	✗	✓	✓	✓
Hebrew BSs and BPs	✗	✗	✓	✗
Classifier languages				
Mandarin BNs	✓	✓	✓	✓

Table 1   Language-specific predictions of the KA for the availability of BNs in standard argument positions in (in)definite singular/plural contexts.

	Singular INDEF.	Singular DEF.	Plural INDEF.	Plural DEF.
Non-classifier languages				
Spanish BSs and BPs	✗	✗	✓	✗
German BSs and BPs	✗	✗	✓	✗
Russian BSs and BPs	✓*	✓	✓	✓
Hindi BSs and BPs	✓*	✓	✓	✓
Hebrew BSs and BPs	✓*	✗	✓	✗
Classifier languages				
Mandarin BNs	✓	✓	✓	✓

Table 2   Language-specific predictions of the PA for the availability of BNs in standard argument positions in (in)definite singular/plural contexts.

To finetune the predictions for the languages in our sample, we follow the assumptions about articles in the KA literature (see (6)). Combining these assumptions with the above predictions leads us to the language-specific predictions in Table 2. The table indicates that BNs are expected to be available when there is no corresponding article and it signals their unavailability exactly for those cases for which the languages do have a corresponding article: Spanish and German singular indefinite and singular/plural definites, and Hebrew singular and plural definites. The cells including an asterisk mark the differences from the predictions of the KA in Table 1.

4   A parallel corpus study

With the predictions of the KA and the PA in place, we can compare them side by side. Table 1 and Table 2 reveal that the two approaches make very similar predictions. For plural contexts, the two approaches are indistinguishable (but see our discussion of scope in Section 1). For singular contexts, we do find several differences, specifically in indefinite contexts in languages for which we have assumed that they lack indefinite articles. In such contexts, the PA predicts Russian, Hindi and Hebrew BSs to appear freely, whereas the KA predicts them to be unavailable. The predicted unavailability in the KA is due to a combination of the type-shift ranking in (4) and the unavailability of the *down* shift for singulars. BNs in Mandarin escape these restrictions in the KA because they start life as non-singular kinds, allowing them to undergo DKP. Given that the predictions of the KA and PA only differ significantly for the singular domain, we will restrict our parallel corpus study to singular definite and indefinite contexts.

Before spelling out our methodology in more detail, we need to get back to the role of pseudo-incorporation that we briefly hinted at in Section 2. As Dayal (2004) already pointed out, the appearance of BNs in indefinite contexts may involve cases of pseudo-incorporation, which would fall outside the scope of the predictions in Table 1 and Table 2. Therefore, one should be mindful that BNs in indefinite contexts do not all necessarily appear in standard argument positions when considering the actual data. For the languages in our sample, pseudo-incorporation has been suggested to play a role in Hindi (Dayal 2004, 2011), Spanish (Dobrovie-Sorin, Bleam & Espinal 2006; Espinal & McNally 2011), Hebrew (Doron 2003) and Mandarin (Huang 2015; Luo 2022). At the same time, little is known about the extent of pseudo-incorporation in (some of) these languages. The current study assumes, in line with Dayal 2011, the view that pseudo-incorporation has limited productivity in those languages that make use of it. Therefore, for our data, we will assume that the explanatory potential of pseudo-incorporation is limited to languages with a restricted distribution of BNs in indefinite contexts.

4.1   Methodology

As indicated in Section 1, we rely on a translation corpus to map out cross-linguistic variation under conditions of maximal comparability. Our corpus consists of the translations of the first chapter of *Harry Potter and the Philosopher's Stone* to Spanish, German, Hebrew, Russian, Hindi, and Mandarin. With the KA and PA leading to diverging predictions for singulars, we use the translations of *a* N_{sg} (N=90) and *the* N_{sg} (N=140) as arguments of verbs and prepositions as proxies for the singular definite and indefinite domains. We compare the distributions of BNs to their main

competitors as they emerge from our data: the indefinite article and the numeral *one* for indefinites and the definite article and demonstrative for definites. Example sentences (7) and (8) illustrate the data in our corpus, presenting two English indefinite contexts, followed by their translations in the languages of our sample.

(7) "She told him over dinner all about Mrs Next Door's problems with her daughter and how Dudley had learnt **a new word** ('Shan't!')."

a. **Spanish**

Mientras comían,   le   informó de los problemas de la   señora Puerta
while   were.eating.3P him informed of the problems   of the Mrs   Door
Contigua con su hija,   y   le   contó que Dudley había aprendido **una**
Next   with her daughter and her told   that Dudley had   learnt   a
nueva frase   ("¡no lo haré!").
new   phrase   not it will.do.1P

b. **German**

Beim Abendessen erzählte sie ihm alles über Frau Nachbarins Probleme mit
at.the dinner   told   she him all   about Mrs Neighbour's problems with
deren Tochter   und dass Dudley **ein neues Wort** gelernt hatte ("pfui").
her   daughter and that Dudley a   new   word learnt   had   ugh

c. **Hebrew**

bə-ʔaruxat   ha-ʔerev   hi sipra l=o   ʔal   kol ha-bəʔajot ʃə-yeʃ
at-meal.POSS the-evening she told to-him about all the-problems that-there.are
la-ɡveret   ha-ʃxena   ʔim ha-bat   ʃela ve-ʔal   ze ʃe-dadli
to/the-woman the-neighbour with the-daughter of-her and-about it that-Dudley
lamad ha-jom **bituj   xadaʃ** (lo   rotse).
learnt the-day expression new   not   want

d. **Russian**

Za obedom ona oxotno spletnčiala, rasskazav misteru   Dursley o   tom,
at lunch   she gladly gossiped   having.told mister.DAT Dursley about that
čto u ix   sosedki   ser'ëznye problemy s   dočer'ju, i   naposledok
that at their neighbour serious   problems with daughter and finally
soobščiv,   čto Dudley vyučil **novoe slovo** "xaččʉ!".
having.informed that Dudley learnt new   word I.wanna

e. **Hindi**

Unho-ne dinner par apne pati   ko bata-ya ki   padosan   ki apni beti
She-ERG dinner on her husband to told-PFV that neighbour of own daughter
ke.sath kya   samasyaye chal rahi hai   aur Dudley-ne **ek naya vakya**
with   what problems   go   PROG be.PRES and Dudley-ERG a   new   sentence
sikh-a   hai   'nahi karu-n-ga'.
learn-PFV be.PRES no   do-FUT-M

f. **Mandarin**

wănfàn zhuō shàng, désīlǐ tàitài xiàng tā jiǎngshù-le línjū jiā de
dinner table on Durseley Mrs to he tell-ASP neighbour family DE
mǔ-nǚ máodùn, hái shuō dálì yòu xué-huì **yí-gè xīn-cí**
mom-daughter conflict also say Dudley again learn-RVC one-CL new-word
("juébù").
 never

(8) "It was on the corner of the street that he noticed the first sign of something peculiar - a cat reading **a map**."

a. **Spanish**

Al llegar a la esquina percibió el primer indicio de que
at.the arrive at the corner noticed.3P the first sign of that
sucedía algo raro: un gato estaba mirando **un plano de la**
was.happening.3P something strange a cat was looking a plan of the
ciudad.
city

b. **German**

An der Straßenecke fiel ihm zum ersten Mal etwas Merkwürdiges auf -
at the street.corner felt him for.the first time something strange PREF
eine Katze, die **eine Straßenkarte** studierte.
a cat that a street.map studied

c. **Hebrew**

rak bə-keren ha-rexov hu hivxin ba-siman ha-riʃon lə-maʃəhu muzar -
only at-corner the-stret he noticed in/the-sign the-first of-something weied -
xatula ʃə-ʔijna bə-**mapa**.
cat.F that-read in-map

d. **Russian**

Tol'ko na uglu ulicy mister Dursley nakonec zametil, čto proisxodit
only on corner street.GEN mister Dursley finally noticed that happens
čto-to strannoe, a zametil on košku, vnimatel'no izučavšuju ležaščuju
something strange and noticed he cat.ACC attentively examining lying
pered nej **kartu**.
in.front.of her map.ACC

e. **Hindi**

Sadak-ke mod par dursley ko pehli ajib chiz dikh-i ek billi, jo
Street-GEN corner on Dursley to first strange thing.F see-PST.F a cat.F who
naksha padh rahi thi.
map read PROG be.PST

f. **Mandarin**

zài jiē-jiǎo   shàng, tā kàn-dào-le   dì-yī-gè   yìcháng-de xìnhào
at street-corner on   he see-RCV-ASP ORD-one-CL peculiar-DE sign
yì-zhī-māo zài   kàn **dìtú**.
one-CL-cat PROG read map

4.2 Results

Figures 1 and 2 summarize the data, the former spelling out the translations of *a* N_{sg}, the latter those of *the* N_{sg}. We present a brief run-through organized per language. Spanish and German by and large come out as languages in which nouns in singular definite contexts require the definite article and nouns in singular indefinite contexts require the indefinite article. Hebrew comes out as a language in which nouns in singular definite contexts require the definite article but typically appear bare in singular indefinite contexts (Doron 2003). The *rest* category in the Hebrew definite domain is bigger than its Spanish and German counterparts. This is due to the frequent use of the construct state (e.g., ʔaruxat ha-ʔerev, 'dinner', in (7c)), which amounts to 44% (15/34) of the *rest* category and 12% (15/127) of all noun phrases in Hebrew definite contexts. Russian relies on BSs in definite and indefinite contexts alike (Seres & Borik 2021). Hindi BSs have a hybrid position: they freely allow for definite readings but appear next to numeral *ek* ('one') N in the indefinite domain (Dayal 2004). Mandarin BNs appear next to numeral *yi*+CL ('one') N in the indefinite domain (Li & Thompson 1989) and demonstratives show an increased use in the definite domain (Jenks 2018; Bremmers et al. 2022; Jiang & Dayal 2023).

4.3 Discussion

Our data are in line with descriptive generalizations from the literature, but when juxtaposed, they reveal challenges for the KA and the PA alike. We argue that both approaches only account for part of the data.

For Spanish and German, both the KA and PA correctly predict the absence of BSs in argument position. The BSs that we do find in the indefinite domain appear after prepositions and – for Spanish – in the object position of HAVE-verbs, in line with claims about pseudo-incorporation in the literature (Dobrovie-Sorin et al. 2006; Espinal & McNally 2011). We conclude that the KA and PA are equally successful in accounting for the Spanish and German facts.

For Hindi, we find that the PA does not make the right predictions but the KA does, *modulo* an important role for pseudo-incorporation. The Hindi data challenges the PA in the sense that the absence of definite and indefinite articles that we assumed with the literature (see 6), leads the PA to predict BSs to freely appear
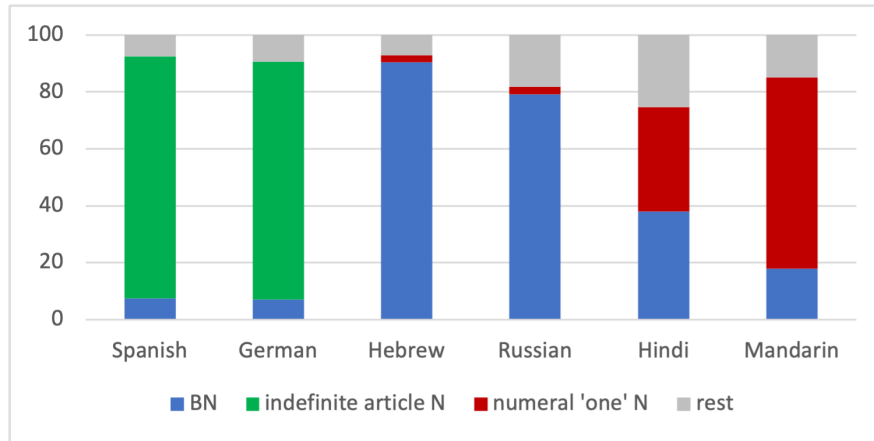
Figure 1 Spanish, German, Hebrew, Russian, Hindi and Mandarin translations of **a + N$_{sg}$** (%) in Chapter 1 of *Harry Potter and the Philosopher's Stone*
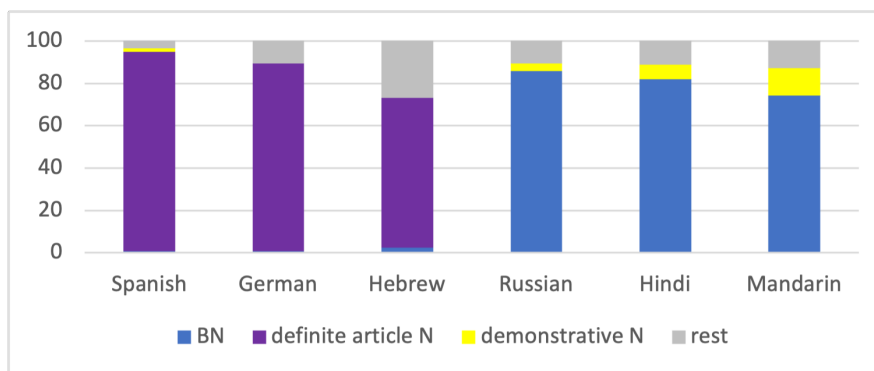
Figure 2 Spanish, German, Hebrew, Russian, Hindi and Mandarin translations of **the + N$_{sg}$** (%) in Chapter 1 of *Harry Potter and the Philosopher's Stone*

in definite and indefinite contexts. The difference in distribution of BSs between these two types of contexts suggests that the PA makes the right predictions for the former but not for the latter, leaving the presence of the numeral in datapoints like (7e) as opposed to its absence in datapoints like (8e) unaccounted for. For the KA, the appearance of BSs in definite contexts is straightforwardly explained and so is the appearance of the numeral in datapoints like (7e). The KA can also account for the absence of the numeral in datapoints like (8e) under the assumption that BSs are allowed in indefinite contexts if they are pseudo-incorporated, as proposed in

Dayal 2004, 2011. We submit that the opposition between (7e) and (8e) is in line with the predicted pattern. Assuming with Le Bruyn, Swart & Zwarts (2016) that pseudo-incorporation is likelier with VO combinations in which the verb doubles a relation that is implicit or explicit in the object noun, the opposition between (7e) and (8e) follows. Indeed, *naksha* ('map') in (8e) has a READ relation as part of its telic quale (Pustejovsky 1995) and this is doubled by the verb *padh* ('read'), predicting the availability of the BN as part of a pseudo-incorporation construction. *Vakya* ('sentence') in (7e) arguably does not come with any implicit *learning* relation in its qualia, making its combination with *sikh* ('learn') unlikely to allow for pseudo-incorporation. Given our current assumptions about articles, the KA but not the PA can account for the distribution of BSs in both definite and indefinite contexts. We conclude that the Hindi data *prima facie* favor the KA over the PA.

For Hebrew and Russian, however, the tables turn, and only the PA straightforwardly makes the right predictions. The absence of definite and indefinite articles in Russian leads the PA to correctly predict BSs to freely appear in definite and indefinite contexts. The KA makes the right predictions in definite contexts but fails to extend its success to indefinite contexts, where it predicts BSs to be unavailable, contrary to fact (see (7d) and (8d)). If we were to analyze all Russian BSs in indefinite contexts as pseudo-incorporated, they would fall outside the scope of the KA's predictions for standard argument positions. However, we discard this theoretical possibility since BSs in Russian seem to lack the restricted use that pseudo-incorporation makes us expect (see Section 4).[2] For Hebrew, the presence of a definite article and the absence of an indefinite article lead the PA to correctly predict BSs to appear freely in indefinite contexts but not in their definite counterparts. The KA, in sharp contrast to the PA's success in capturing the Hebrew data, is challenged by the availability of BSs in indefinite contexts (see (7c) and (8c)), and a pseudo-incorporation route lacks empirical support in Hebrew as it does in Russian. We conclude that the Hebrew and Russian data favor the PA over the KA.

For Mandarin, the predictions of the PA and KA are more in line with our BN data in definite than in indefinite contexts. Under the assumption that Mandarin does not have articles, the PA makes the prediction that BNs should be equally acceptable in indefinite and in definite contexts, contrary to fact. The KA faces the same problem: given that Situation Restriction and DKP are expected to be equally available for Mandarin BNs, the KA fails to predict the marked difference we find in their distribution between definite and indefinite contexts. For both approaches, the relatively high proportion of demonstratives in definite contexts (13%) especially in comparison to Russian (4%) also comes as a surprise. We conclude that the Mandarin data are problematic for both the KA and the PA.

---

2 See Mueller-Reichau (2015) for a detailed discussion of pseudo-incorporation in Russian.

Taking stock, we have argued that the KA and PA make the right predictions for Spanish and German. For Hebrew, Russian and Hindi, we have argued that both approaches make the right predictions for BNs in definite contexts but only successfully account for BNs in indefinite contexts for a subset of the languages: as it stands, the PA makes the right predictions for Hebrew and Russian but not for Hindi, whereas the KA makes the right predictions for Hindi but not for Hebrew or Russian. Finally, for Mandarin, we find that both approaches have trouble accounting for BNs in both indefinite and definite contexts.

4.4   The explanatory potential of the PA and the KA

With the predictions of neither approach being fully borne out, the question that imposes itself is whether we can tweak either or both to reach empirical adequacy. We start with definite contexts and then move to indefinite ones.

For definite contexts, the only language that leads to problems for the PA and KA is Mandarin. The recent literature is converging on the idea that Mandarin BNs and demonstratives are in complementary distribution even though the details remain to be worked out (see, e.g., Jenks 2018; Bremmers et al. 2022; Simpson & Wu 2022). We are confident that both approaches can be extended to cover the data. For the PA, this would involve a refinement of the Blocking Principle. For the KA, something akin to the Blocking Principle would need to be developed to model the interaction between the demonstrative and BNs at the level of SR.

For the indefinite data, we first discuss possible extensions of the PA and then move to the KA. We argue that the PA can straightforwardly be extended to cover the totality of the data. To do so, our first step is to change our assumptions about the cross-linguistic inventory of indefinite articles: we originally followed the KA literature in assuming that Hebrew, Russian, Hindi, and Mandarin lack indefinite articles. Our proposal is to change this assumption for Hindi and Mandarin and to take Hindi *ek* and Mandarin *yi*+CL to function as indefinite articles. With this assumption in place, the PA predicts the Blocking Principle to kick in and block Hindi BSs and Mandarin BNs from appearing in singular indefinite contexts.

Next, we assume with Dayal (2004, 2011) that the availability of Hindi BSs in singular indefinite contexts relies on pseudo-incorporation and we extend this assumption to Mandarin BNs (see also Huang 2015; Luo 2022). It follows from this position that the PA no longer predicts Hindi BSs and Mandarin BNs to be excluded from singular indefinite contexts, but rather that their use should be restricted. This prediction is in line with the tendencies we find in the frequency data in Figures 1 and 2. It also straightforwardly explains why the presence/absence of the Hindi numeral in (7e) and (8e) neatly patterns with the presence/absence of the Mandarin numeral in (7f) and (8f).

Moving to the KA, we argue that the extensions we proposed for the PA do not affect the predictions of the KA and explain why we think that the latter cannot be extended to cover the totality of the data. As for the extensions we proposed for the PA, the assumption that Hindi *ek* functions as an indefinite article merely doubles the restriction in the KA on BSs in singular indefinite contexts that follows from the ranking of the *iota* shift above the existential shift. The assumption that Mandarin *yi*+CL also functions as an indefinite article furthermore has no impact on the predictions of the KA given that DKP is not part of the regular set of type-shifts and the Blocking Principle does not apply to it. We conclude that the extensions we proposed for the PA do not have a direct impact on the predictions of the KA.

As for the difficulties we see to extend the KA to cover the totality of the data, we think it would be feasible to cover the Hebrew data but neither the Russian nor the Mandarin data. For Hebrew, one could work out an analysis in which the existential type-shift becomes available to BSs because the *iota* and the *down* shifts are independently blocked. A similar escape route is not available in Russian, where the availability of BSs in definite contexts shows that the *iota* shift is not blocked. To cover the unrestricted availability of Russian BSs in indefinite contexts, the only path we see for the KA is to abandon the assumption that the *iota* shift is ranked above the existential shift. However, this would mean that a fundamental insight of the approach must be abandoned. As for the restriction we find on the use of BNs in indefinite contexts in Mandarin, we see no easy way to account for it unless we assume DKP is simply not available, a yet more problematic move than reconsidering the type-shift ranking. The problem is that Mandarin BNs are unrestricted in plural indefinite contexts (Liu, Patil, Seres, Borik & Le Bruyn to appear), strongly suggesting that the restrictions in singular indefinite contexts really come from an interaction between BNs and *yi*+CL. However, the KA–in its current version–has no level at which DKP could be made to interact with the regular type-shifting operations that are involved in the use of the numeral.

Summarizing the discussion, we have argued that both the PA and the KA are likely to be able to account for the use of BNs in definite contexts but that only the PA can easily be extended so as to cover the full set of indefinite data we found. The main challenges for the KA lie in (i) the clearcut opposition between Russian and Hindi BNs in indefinite contexts and (ii) the restricted distribution of BNs in Mandarin. For the PA, we have argued that it can be extended to cover the totality of the data if we assume–contrary to the consensus in the KA literature–that the numeral *one* in Hindi and Mandarin functions as an indefinite article and if we furthermore assume that pseudo-incorporation can account for the indefinite uses of BNs in these languages in singular indefinite contexts. Obviously, the latter assumption calls for a follow-up study in which BNs in indefinite contexts are studied through the lens of pseudo-incorporation, paying close attention to the diverging

distribution of BSs/BNs in the languages for which we take pseudo-incorporation to be at play (Spanish, Hindi, and Mandarin).

4.5 Two notes on methodology

We end our discussion with two methodological notes regarding parallel corpus research in general and the level of parallelism we have pursued in the current study. Le Bruyn & de Swart (submitted) take parallel corpus research to be a valuable addition to the cross-linguistic semanticist's toolbox, but one that ultimately requires replication and triangulation to accumulate critical mass across studies. The current paper is best considered in this spirit as a proof of concept. As such, the data it brings to bear on the availability of BNs should be taken with a grain of salt, and the conclusions that we draw from it handled with care. For future work, we intend to run the same analysis on the same HP chapter, translated by different translators for a subset of the languages in our sample, namely, Russian, Hindi, and Mandarin, which all lead to diverging predictions for the PA and the KA. For these languages, the first HP volume happens to have at least two official translations, allowing us to assess how representative our data is for these languages. We also wish to extend the number of languages we examine and refer the interested reader to Borik, Le Bruyn, Liu & Seres (to appear) for a closer look at parallel data from Polish and Macedonian, the former patterning with Russian, the latter patterning with Hebrew in the definite domain but exhibiting a more extensive use of the numeral *one* in the indefinite domain.

Regarding the level of parallelism we have pursued in this comparative study, the attentive reader may have noticed that we present our frequency data per expression per language but not per context. The upshot of this is that one cannot directly evaluate how the contexts in which an expression α in one language appears relate to the contexts in which an expression β in another language appears. The choice not to go for parallelism at the level of contexts is inspired by the size of our corpus: every translation brings in a limited number of idiosyncratic choices, but these add up with every language we add, making it hard to discern the bigger patterns in our fairly small dataset if we present the data per expression per language per context. In future work, we will be adding more chapters of the same translations and pursue parallelism at the level of contexts in the way we have done in other studies under the *Translation Mining* approach (see, e.g., Bremmers et al. 2022; Klis et al. 2022). This will allow us–in particular–to really probe the variation in pseudo-incorporation that emerges from our data.

5 Conclusion

In this paper, we adopted a translation corpus approach to reassess the empirical coverage of two closely related theories of argument formation: Chierchia's Kinds Approach (KA) (Chierchia 1998; Dayal 2004; Jiang 2020) and Krifka's (2003) Properties Approach (PA). Given that both theories make the same predictions for plural contexts, we focused on singular contexts. Our corpus consisted of the translations of the first chapter of *Harry Potter and the Philosopher's Stone* to Spanish, German, Russian, Mandarin, Hindi, and Hebrew.

We argued that both the KA and the PA make the right predictions for singular definite contexts but that only the PA can be extended to account for the patterns we found for singular indefinite contexts. To derive these patterns in the PA, we hypothesized that the numeral *one* in Hindi and Mandarin functions as an indefinite article and that the BNs that appear in singular indefinite contexts in the two languages are to be accounted for using pseudo-incorporation. The challenges we identified for the KA lie in the unconstrained distribution of BNs in singular indefinite contexts in Russian and in the constrained distribution of BNs in the same contexts in Mandarin. The first challenge calls the KA's type-shift ranking into question, the second the status of DKP.

Relevant follow-up research that we identified includes replication of the findings on the basis of a second set of official translations of the same corpus for Russian, Hindi and Mandarin as well as an extension of the current corpus to pursue parallelism at the level of contexts and probe the variation we find in more detail. Special attention would need to be paid to the varying extent of pseudo-incorporation, especially in Spanish, Hindi and Mandarin.

References

Borik, Olga, Bert Le Bruyn, Jianan Liu & Daria Seres. to appear. Bare nouns in Slavic and beyond. In Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.), *Advances in Formal Slavic Linguistics 2022* Open Slavic Linguistics, Berlin: Language Science Press.

Bremmers, David, Jianan Liu, Martijn van der Klis & Bert Le Bruyn. 2022. Translation mining: Definiteness across languages (a reply to Jenks 2018). *Linguistic Inquiry* 53(4). 735–752. doi:10.1162/ling_a_00423.

Carlson, Greg N. 1977. A unified analysis of the English bare plural. *Linguistics and philosophy* 1(3). 413–457. doi:10.1007/BF00353456.

Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural language semantics* 6(4). 339–405. doi:10.1023/a:1008324218506.

Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174. 99–149. doi:10.1007/s11229-009-9686-6.

Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and philosophy* 27. 393–450. doi:10.1023/B:LING.0000024420.80324.67.

Dayal, Veneeta. 2011. Hindi pseudo-incorporation. *Natural Language & Linguistic Theory* 29(1). 123–167. doi:10.1007/s11049-011-9118-4.

Dobrovie-Sorin, Carmen, Tonia Bleam & M. Teresa Espinal. 2006. Bare nouns, number and types of incorporation. In Svetlana Vogeleer & Liliane Tasmowski (eds.), *Non-definiteness and plurality* Linguistik Aktuell/Linguistics Today 95, 51–79. John Benjamins Amsterdam. doi:10.1075/la.95.04dob.

Doron, Edit. 2003. Bare singular reference to kinds. In *Semantics and Linguistic Theory*, vol. 13, 73–90. doi:10.3765/salt.v0i0.2881.

Espinal, M. Teresa & Louise McNally. 2011. Bare nominals and incorporating verbs in Spanish and Catalan. *Journal of Linguistics* 47(1). 87–128. doi:10.1017/S0022226710000228.

Gehrke, Berit. 2022. Differences between Russian and Czech in the use of aspect in narrative discourse and factual contexts. *Languages* 7(2). 155. doi:10.3390/languages7020155.

Huang, C-T James. 2015. On syntactic analyticity and parametric theory. *Chinese syntax in a cross-linguistic perspective* 1. 48. doi:10.1093/acprof:oso/9780199945658.003.0001.

Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49(3). 501–536. doi:10.1162/ling_a_00280.

Jiang, Li & Veneeta Dayal. 2023. The puzzle of anaphoric bare nouns in Mandarin: A counterpoint to *Index!* *Linguistic inquiry* 54(1). 147–167. doi:10.1162/ling_a_00433.

Jiang, Li Julie. 2020. *Nominal arguments and language variation*. Oxford University Press. doi:10.1162/ling_a_00280.

Klis, Martijn van der, Bert Le Bruyn & Henriëtte de Swart. 2022. A multilingual corpus study of the competition between PAST and PERFECT in narrative discourse. *Journal of Linguistics* 58(2). 423–457. doi:10.1017/S0022226721000244.

Krifka, Manfred. 2003. Bare NPs: Kind-referring, indefinites, both, or neither? In *Semantics and linguistic theory*, vol. 13, 180–203. doi:10.3765/salt.v13i0.2880.

Le Bruyn, Bert, Martín Fuchs, Martijn van der Klis, Jianan Liu, Chou Mo, Jos Tellings & Henriëtte de Swart. 2022. Parallel corpus research and target language representativeness: The contrastive, typological, and translation mining traditions. *Languages* 7(3). 176. doi:10.3390/languages7030176.

Le Bruyn, Bert, Martín van der Kliss & Henriëtte de Swart. to appear. Variation and stability: the HAVE-PERFECT and the tense-aspect grammar of Western

European languages. In Frank Brisard, Astrid ded Wit, Carol Madden, Michael Meeuwis & Adedline Patard (eds.), *Beyond Time*, Oxford University Press.

Le Bruyn, Bert & Henriëtte de Swart. submitted. Cross-linguistic research, parallel corpora, and replication in the translation mining tradition. 23rd Annual International Conference of the English Departmen (AICED 2022).

Le Bruyn, Bert & Henriëtte de Swart. 2022. Exceptional wide scope of bare nominals. *Semantics and Pragmatics* 15. doi:10.3765/sp.15.7.

Le Bruyn, Bert, Henriëtte de Swart & Joost Zwarts. 2016. From HAVE to HAVE-verbs: Relations and incorporation. *Lingua* 182. 49–68. doi:10.1016/j.lingua.2016.04.002.

Li, Charles N & Sandra A Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. University of California Press.

Liu, Jianan, Shravani Patil, Daria Seres, Olga Borik & Bert Le Bruyn. to appear. 'Articleless' languages are not created equal. Sinn und Bedeuting (SuB 27).

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of n-movement in syntax and logical form. *Linguistic inquiry* 609–665.

Luo, Qiongpeng. 2022. Bare nouns, incorporation, and event kinds in Mandarin Chinese. *Journal of East Asian Linguistics* 31(2). 221–263. doi:10.1007/s10831-022-09239-1.

Mueller-Reichau, Olav. 2015. Pseudo-incorporation in Russian? Aspectual competition and bare singular interpretation. In Olga Borik & Berit Gehrke (eds.), *The Syntax and Semantics of Pseudo-Incorporation*, 262–295. Leiden: Brill. doi:10.1163/9789004291089_009.

Mulder, Gijs, Gert-Jan Schoenmakers, Olaf Hoenselaar & Helen de Hoop. 2022. Tense and aspect in a Spanish literary work and its translations. *Languages* 7(3). 217. doi:10.3390/languages7030217.

Partee, Barbara H. 1987. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stockhof (eds.), *Studies in discourse intepretation and the theory of generalized quantifiers*, 115–143. Amsterdam: Foris. doi:10.1515/9783112420027-006.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, Massachusetts: MIT Press. doi:10.7551/mitpress/3225.001.0001.

Seres, Daria & Olga Borik. 2021. Definiteness in the absence of uniqueness: the case of Russian. In Andreas Blümel, Jovana Gajić, Ljudmila Geist, Uwe Junghanns & Hagen Pitsch (eds.), *Advances in Formal Slavic Linguistics 2018* Open Slavic Linguistics, 339–363. Berlin: Language Science Press. doi:10.5281/zenodo.5155544.

Simpson, Andrew & Zoe Wu. 2022. Constraints on the representation of anaphoric definiteness in Mandarin Chinese. In Andrew Simpson (ed.), *New Explorations in Chinese Theoretical Syntax: Studies in honor of Yen-Hui Audrey Li* Linguis-

Jianan Liu
Institute for Language Sciences
Utrecht University
Trans 10, 3512 JK Utrecht
The Netherlands
j.liu4@uu.nl

Hagay Schurr
Ph.D. Program in Linguistics
City University of New York
The Graduate Center
365 Fifth Avenue, Room 7407
New York, New York 10016
United States of America
hschurr@gradcenter.cuny.edu

Olga Borik
Departamento de Filologías Extranjeras y sus Lingüísticas
Universidad Nacional de Educación a Distancia
Paseo Senda del Rey 7, 28040 Madrid
Spain
oborik@flog.uned.es

Shravani Patil
Tilburg University
Warandelaan 2, 5037 AB Tilburg
The Netherlands
s.s.patil@tilburguniversity.edu

Daria Seres
515 Institut für Slawistik
Universität Graz
Merangasse 70/I, 8010 Graz
Austria
daria.seres-guillen@uni-graz.at

Bert Le Bruyn
Institute for Language Sciences
Utrecht University
Trans 10, 3512 JK Utrecht
The Netherlands
b.s.w.lebruyn@uu.nl