The theory of argument formation: between kinds and properties

Chierchia (1998) developed a cross-linguistic extension to Carlson's kinds approach to bare nouns (BNs), producing the most influential theory of argument formation to date. However, recent developments within and outside the neo-Carlsonian framework level the playing field with its closest competitor: a properties approach that takes over Chierchia's generalized blocking principle but leaves the relevant type-shifts unranked and takes BNs to start life as type <e,t>. Within this properties approach (PA), Krifka (2004) argues that a locality constraint on type-shifts overall has the same narrow scope effect on BNs as Chierchia's Derived Kind Predication (DKP). Le Bruyn & de Swart (2022) identify the scope of Dutch scrambled bare plurals as a crucial testing ground and empirically rule in favor of the PA. Within the kinds approach (KA) itself, cognitive and linguistic considerations have pushed researchers to derive the need for classifiers (CL) in CL languages from a mere type-clash (Jiang 2020), count and mass kinds being of type e and numerals requiring <e,t> input, instead of relying on the structure of kinds and the way they map to properties. Although this type-clash derives the need for CLs semantically, it lacks the deeper semantic motivation from Chierchia (1998), allowing its effect to be mimicked in various ways compatible with the PA. Since the latter makes better predictions about scope and as the need for CLs no longer follows from a deeper semantics of kinds, a re-evaluation of the two approaches imposes itself. In this paper, we compare the distributions of BNs in a broad sample of languages, identify new challenges and assess the explanatory potential of the two approaches.

Methodology | We adopt *Translation Mining* as our method (Bremmers et al. 2022; van der Klis et al. 2022), using translations of the same text into different languages to map out cross-linguistic variation. Our corpus consists of the translations of the first chapter of *Harry Potter & the Philosopher's Stone* to Spanish, German, Hebrew, Russian, Hindi, and Mandarin. With the two approaches leading to diverging predictions mainly for singulars, we focus on the translations of *a* N_{sg} (N=90) and *the* N_{sg} (N=140) and use them as proxies for the singular definite and indefinite domain. We compare the distributions of BNs to the main competitors emerging from our data: the indefinite article and the numeral 'one' for indefinites and the definite article and demonstrative for definites.

Results | Graphs 1 and 2 (p.3) summarize the data. **Spanish** and **German** come out as languages with definite and indefinite articles, **Hebrew** as a language with a definite (*ha*) but without an indefinite article (Doron 2004). The Hebrew 'rest' category in the definite domain is relatively big due to construct states. **Russian** relies on BNs in the definite and indefinite domain and counts as an articleless language (Seres & Borik 2021). **Hindi** BNs have a hybrid position: they freely allow for definite readings but appear next to the numeral *ek* in the indefinite domain (Dayal 2004). **Mandarin** BNs appear next to the numeral *yi*+CL in the indefinite domain (Li & Thompson 1989) and show an increasing use of demonstratives in the definite domain (Bremmers et al. 2022; Dayal & Jiang 2022).
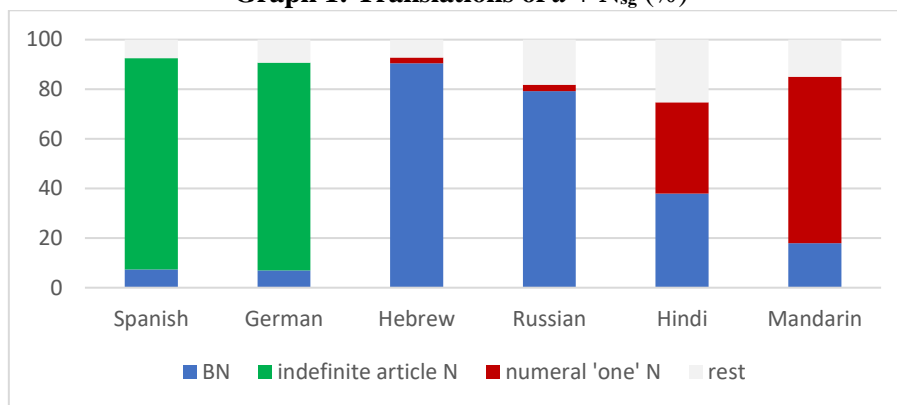
Discussion | Our data are in line with descriptive generalizations from the literature, but their juxtaposition is new and reveals the adequacy of the KA & PA for some languages but not for all. For **Spanish** and **German**, the KA & PA both correctly predict the existence of articles to block BNs from appearing in argument position. The BNs that we do find in the indefinite domain appear after prepositions and in the object position of HAVE-verbs (Espinal & McNally 2010). For **Hebrew**, the KA & PA also make the right predictions. On

the PA, *ha* can be taken to block BNs in the definite domain and, in the absence of an indefinite article, BNs are correctly predicted to freely appear in the indefinite domain. On the KA, *ha* blocks BNs from appearing in the definite domain and the ∃-shift is available to BNs because $^\cap$ is undefined for singulars and *ha* blocks ι. We conclude that the KA & PA can account for the Spanish, German, and Hebrew data. The **Russian**, **Hindi** and **Mandarin** data do pose challenges in the indefinite domain for both approaches. For the PA, the main challenge lies in Hindi and Mandarin: there seems to be blocking of BNs in the indefinite domain, but it is partial. For the KA, the Mandarin pattern is not predicted: Mandarin BNs are assumed to start life as (non-singular) kinds and are predicted to freely allow for indefinite readings through DKP (Jiang 2020), especially because none of the indefinites in our corpus takes wide scope. For Russian and Hindi, Chierchia (1998) and Dayal (2004) can each account for one but not for the other. On the former ($^\cap$>ι,∃), the Russian indefinite data follow but the Hindi ones do not, on the latter ($^\cap$,ι>∃), the Hindi indefinite data follow – with the *proviso* that indefinite BNs are incorporated – but the Russian ones do not. We conclude that neither the PA nor the KA provide a robust theory of argument formation that can derive the Russian, Hindi, and Mandarin data in parallel.
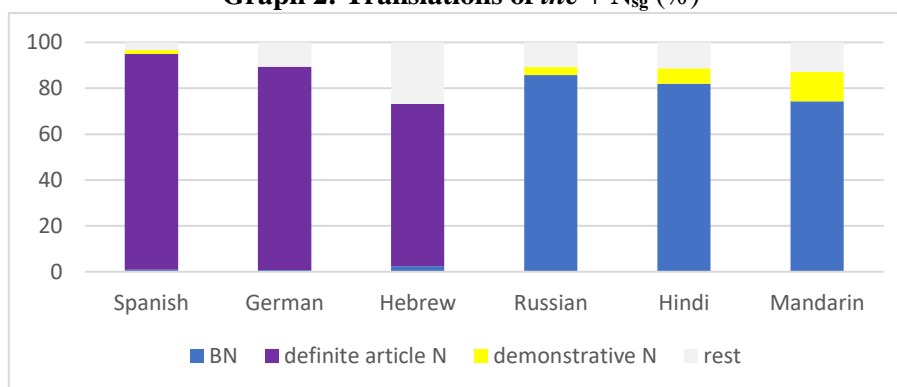
Analysis | We argue that the PA but not the KA can straightforwardly be extended to cover the Russian, Hindi, and Mandarin data. For our extension to the PA, we make two assumptions: (i) Hindi *ek* and Mandarin *yi*+CL function as articles; (ii) Hindi and Mandarin but not Russian allow for pseudo-incorporation (PI) (Dayal 2011; Luo 2022; Mueller-Reichau 2015). With these assumptions in place, the basic patterns we find for Russian, Hindi, and Mandarin follow on the PA. Russian 'one' is not an article, hence it does not block the ∃-shift, accounting for the free use of indefinite BNs. The complementary distribution of BNs and *ek/yi*+CL in Hindi and Mandarin follows from the assumption that the numerals function as articles but are semantically redundant in Verb-Object-combinations that allow for PI, clearing the way for BNs there. Crucially, the extension we propose derives the generalizations only when built into the PA, not when built into the KA. Indeed, given that DKP is an operation that goes straight from type e to type <<e,t>,t>, the KA cannot model an interaction between DKP and indefinite articles, which would leave the pattern in the Mandarin data unaccounted for. For the standard variant of the KA that ranks ι over ∃, the non-existence of an indefinite article and PI in Russian moreover would not overturn the predicted absence of BNs in the indefinite domain. A clear prediction of our PA-based proposal is that all occurrences of BNs in Hindi and Mandarin are cases of PI. To verify this, we need a theory of PI that defines it on independent grounds. We argue that Le Bruyn et al. (2016) propose such a theory, laying the foundation for a competition analysis between indefinite articles and PI.

Conclusion | The kinds approach (KA) to BNs is the most influential theory of argument formation to date but recent developments put a re-assessment of its relationship with the properties approach (PA) high on the agenda. We found that – when analyzed in parallel – Russian, Hindi, and Mandarin BNs do not receive a satisfactory analysis on the KA nor on the PA. We argued that the PA but not the KA can straightforwardly be extended by analyzing Hindi and Mandarin *ek/yi*+CL as indefinite articles and by developing a competition analysis between articles and BNs in pseudo-incorporation contexts. We conclude that parallel data provide new ways of looking at cross-linguistic variation and that this variation reveals a clear advantage of the PA over the KA.

Graph 1: Translations of *a* + N$_{sg}$ (%)

Legend: BN (blue), indefinite article N (green), numeral 'one' N (dark red), rest (light grey). Categories on x-axis: Spanish, German, Hebrew, Russian, Hindi, Mandarin. Y-axis from 0 to 100.

Graph 2: Translations of *the* + N$_{sg}$ (%)

Legend: BN (blue), definite article N (purple), demonstrative N (yellow), rest (light grey). Categories on x-axis: Spanish, German, Hebrew, Russian, Hindi, Mandarin. Y-axis from 0 to 100.

Selected references

Bremmers, D., Liu, J., Klis, M. van der, & Le Bruyn, B. (2022). Translation mining: definiteness across languages (a reply to Jenks 2018). *Linguistic Inquiry*, *53*(4), 735-752.

Dayal, V., & Jiang, L. J. (2020). The Puzzle of Anaphoric Bare Nouns in Mandarin: A Counterpoint to Index!. *Linguistic Inquiry*, 1-42.

Jiang, L. J. (2020). *Nominal arguments and language variation*. Oxford University Press.

Klis, M. van der, Le Bruyn, B., & Swart, H. de (2022). A multilingual corpus study of the competition between past and perfect in narrative discourse. *Journal of Linguistics*, *58*, 423-457.

Krifka, M. (2004). Bare NPs: kind-referring, indefinites, both, or neither?. In *Semantics and linguistic theory* (Vol. 13, pp. 180-203).

Le Bruyn, B., & Swart, H. de (2022). Exceptional wide scope of bare nominals. *Semantics and Pragmatics*, *15*(7).

Le Bruyn, B., Swart, H. de, & Zwarts, J. (2016). From HAVE to HAVE-verbs: Relations and incorporation. *Lingua*, *182*, 49-68.

Luo, Q. (2022). Bare nouns, incorporation, and event kinds in Mandarin Chinese. *Journal of East Asian Linguistics*, *31*, 221-263.

Mueller-Reichau, O. (2015). Pseudo-incorporation in Russian? Aspectual competition and bare singular interpretation. In *The syntax and semantics of pseudo-incorporation* (pp. 262-295). Brill.

Seres, D., & Borik, O. (2018). Definiteness in the absence of uniqueness: The case of Russian. *Advances in formal Slavic linguistics*, 339-363.