Generics Across Languages:
first impressions of developing and using a cross-linguistic toolkit

Background. All languages are thought to have ways to express generalisations over individuals like 'Tigers have stripes' or 'A potato contains vitamin C' (Krifka et al. 1995). These expressions, termed 'generics', inform our understanding of the cognitive capacities of our ability to go beyond particular experiences and communicate about the nature of our world. Generics, present in everyday conversation, convey characteristic or striking properties and express knowledge, including beliefs and stereotypes (e.g., 'Lawyers are greedy'). Linguists have explored their fundamental yet complex nature since the 1970s (Lawler, 1972, 1973; Carlson 1977), while generics have recently gained attention from psychologists and philosophers (Leslie 2007, 2008, Gelman 2008; Sterken 2015; Nickel 2016, among others). Generics have been claimed to be 'defaults' in children and adults (Leslie et al. 2011; Gelman 2008), yet they are not as well understood across human languages as other linguistic forms.

The focus on generics aligns with OASIS purpose and credo, since generics can be seen as one of the building blocks of human cognition, as they reflect our conceptualisation of categories and their properties and the organisation of our experience of the world. This project addresses how these building blocks are expressed with the hope to elucidate the relationship of the formal semantic ontology to grammar and cognition. Furthermore, it incorporates insights from formal linguistics and semantic fieldwork elicitation techniques as well as from cognitive psychology fostering intra- and interdisciplinary communication and examines less-familiar languages broadening the evidence base against which claims about language and cognition are made.

Cross-linguistic research on generics has primarily focused on a few languages (Chierchia 1998; Dayal 2004; Krifka 2004; Behrens 2005). Exceptions going beyond Indo-European include Dahl (1985) and Gerstner-Link's thesis (1995), which lack in consistency and accessibility. Although these studies proposed several key language universals on generics (Dahl 1995: (a) no generic article or marker across languages, (b) minimal tense-aspect marking tendency; Gerstner-Link 1998: (c) if a language grammatically codes definiteness, generic noun phrases are preferably coded as definite, (d) if a language has optional nominal number marking, generics are number neutral, if marking is obligatory, generics are singular or unmarked), there has been scarce literature addressing them and some of the evidence behind them has been recently challenged (Filip, to appear). Furthermore, they were limited in their cross-linguistic scope and phenomenological breadth, because these studies were often based on limited datasets (e.g., researchers had to work on the data available for a given language; the data across languages was not gathered using the same materials and methods). Against this background, the goal of this project is to design a toolkit that can be used across languages for a systematic exploration of generics.

This project addresses the research gap and the shortcomings of previous work by adopting and adapting approaches developed in recent cross-linguistic semantics projects e.g., on quantifiers (Keenan and Paperno 2012), count/mass nouns (Lima and Rothstein 2020), and logical vocabularies including clause-embedding predicates and modals (Uegaki et al. 2024), underscoring the importance of incorporating less-represented languages in theoretical discourse (see Nevins 2022).

Generics Across Languages Toolkit. We are developing and piloting the toolkit, combining insights from formal linguistics and semantic fieldwork elicitation techniques. In Phase 1 of the pilot, the toolkit consisted of 5 storyboards (i.e., short stories supported by pictures, Burton and Matthewson 2015) that included 41 target generic sentences that focused on a different ontological category, including natural and social kinds ('banana', 'lion', 'mud', 'Canadian', 'doctor') and targeting at least five different readings (kind, subkind, characterizing/generic, habitual, and episodic; see examples on pages 2-3), drawing from the linguistics as well as the psychology literature (Krifka et al. 1995; Mari et al. 2013; Prasada et al. 2013). Data in 6 languages were collected: Greek, Hungarian, Japanese, Kinyarwanda, Mongolian and Tshiluba, spanning familiar languages from the generics literature as well as languages with no generics literature. An accompanying administration handbook

explained the reading each sentence was targeting and provided context to aid the consultant target that reading. The toolkit was completed by expert linguist consultants via introspection or via data elicitation from native speakers of the language. The task was to provide a translation of the target sentences and/or to re-tell the story (see sample materials in Figure 1). In Phase 2 of the pilot, the refined toolkit consisted of 5 storyboards ('banana', 'lion', 'Canadian', 'car', 'water') eliciting 41 target generic sentences. Data in 18 languages were collected: Anaang, (Palestinian) Arabic, Basque, Catalan Sign Language, (Mandarian) Chinese, Czech, Finnish, French, Greek, Hebrew, Hungarian, Japanese, Kinyarwanda, Mbya Guarani, Mongolian, Tshiluba, Turkish and Vietnamese.

Analysis. The analysis plan includes the following steps: examining (a) the forms used to express genericity, (b) the (un)markedness/(non-)default character of generic expressions, (c) their multifunctionality, (d) evidence for potential universal restrictions, and (e) implicational restrictions/preferences. Preliminary analysis of the pilot data suggests that none of the languages of the sample includes an article or marker exclusively used with generic noun phrases as per the proposed language universal (a) above. Additional analyses will address the remaining four universals. Differences between the different readings (kind, subkind, characterizing/generic, habitual, and episodic) have been identified as well as differences between animal, artefact and social kinds. Just as a brief illustration, in Mandarin Chinese indefinites can have a subkind reading, but not a kind or generic reading, whereas in Mongolian bare singulars are used in kind/generic readings across different kinds, while bare plurals are preferred for social kinds.
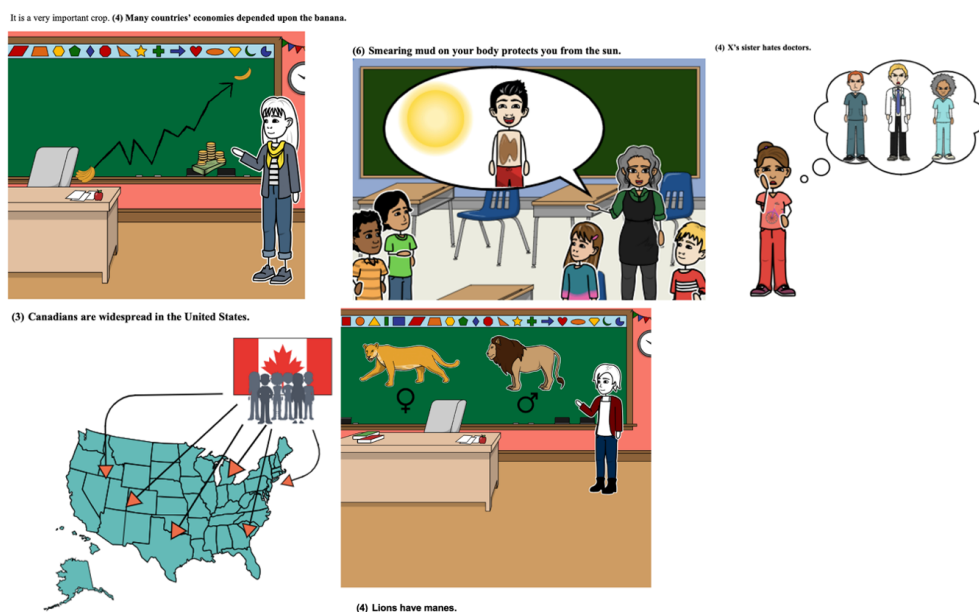

Figure 1. Phase 1 pilot toolkit sample materials.

Conclusion. The toolkit we constructed systematically examines generic sentences with different kinds of nouns and is able to throw light on generalisations that depend on the nature of the noun (count or mass) and/or the kind (natural or social). These first generalisations will be examined against the proposed universals discussed above as well as in light of other well researched phenomena in the literature (e.g. well-defined kind restriction), increasing the evidence base for how generic expressions are expressed and enriching our understanding of generics more broadly.

Selected references. Burton, S., and Matthewson, L. (2015). Targeted Construction Storyboards in Semantic Fieldwork', in M. R. Bochnak, and L. Matthewson (eds), Methodologies in Semantic Fieldwork, New York. **Dahl, O. (1985).** *Tense and Aspect Systems.* Oxford: Blackwell. **Gerstner-Link, C. (1995)**. *Über Generizität. Generische Nominalausdrücke in singulären und generellen Aussagen. München: Wilhelm Fink Verlag.* **Krifka et al. (1995).** Genericity: An introduction. In G. Carlson and F.J.Pelletier (Eds.) The Generic Book, Chicago: CUP, 1-125. **Mari et al. (2013)** Genericity. OUP.