Restrictions on the ontological category of indefinite pronoun series in the languages of Europe

Indefinite pronouns usually consist of two parts: a stem expressing ontological category (PERSON, MANNER, TIME, etc.) and an indefiniteness marker. A general term for indefinite pronouns derived with the same indefiniteness marker basing on different ontological stems is series (e.g., the English system includes *some*-series, *any*-series, etc.). However, the forming of indefinite pronouns may be restricted: not all the logically possible combinations of ontological categories and indefiniteness markers are represented in fact. E.g., the Russian *koe*-series lack the AMOUNT, REASON and DETERMINER pronominal forms: *koe-skol’ko* (INDF-how.many), *koe-počemu* (INDF-why) and *koe-kotoryj* (INDF-which), despite the fact that it is possible to derive indefinite pronouns with other indefiniteness markers from these interrogative stems, cf. *nibud’*-indefinites: *skol’ko-nibud’* (how.many-INDF) ‘any amount’, *počemu-nibud’* (why-INDF) ‘for any reason’, *kotoryj-nibud’* (which-INDF) ‘any (from a set)’.

Hasepalm (1997) claims that the number of absent forms and their distribution over different ontological categories may be accidental. To check this claim, I collected the data about the gaps in indefinite pronoun systems of 21 European languages. The data was partially taken from the 40-language sample collected by Haspelmath (1997: 244–317) and has been extended and re-checked with native speakers so that we know which of the logically possible forms are strictly prohibited. Whereas the number of ontological categories is under discussion (Jackendoff 1990; Haspelmath 1997; Award 2001; Hengeveld, Mackenzie 2008), the following ‘extended’ set of ontological categories has been considered: PERSON, THING, PLACE, TIME, MANNER, AMOUNT, PROPERTY, DETERMINER, REASON, POSSESSION. Then the correlations between mean percentage of absent pronominal forms and different factors that could influence their absence have been measured. The quantitative analysis provided the following results:

1. If a language has both interrogative-based and non-interrogative-based indefinite pronouns, then interrogative-based forms are less likely to be absent. Languages that have only interrogative-based indefinites tend to have fewer absent forms; agglutinative languages usually do not have any restrictions on forming indefinite pronouns at all.

2. The semantics of indefinite pronoun series does not influence significantly neither the frequency, nor the distribution of the absent pronominal forms. Free-choice indefinites have in general fewer gaps, but our data shows that it is related to the fact that many of them are sluicing-based and thus less grammaticalized.

3. The absence of a non-derived AMOUNT or POSSESSION-interrogative stem triggers the absence of a corresponding indefinite pronoun. In contrast, languages that do not have a non-derived REASON-interrogative (that is, a case form of THING-interrogative or an adpositional construction is used instead) have fewer restrictions on forming REASON-indefinites, while languages that have special REASON-interrogatives usually do not derive any REASON-indefinites from them (e.g., English *why* and no indefinites from it). Moreover, *why*-indefinites are strikingly different from other indefinites in many ways. The PCA-analysis shows that the REASON category (the *reason_ratio* vector on the plot) is not the only deviating vector, but it is not close to any other category (the principle component is the mean percentage of absent forms within different ontological categories), see Figure 1. Considering this and the extremely high mean percentage of absent REASON-indefinites in indefinite pronoun systems (0.65 averaged over the sample), one can conclude that REASON is a ‘peripheral’ ontological category. This conclusion meets our expectations regarding the numerous evidences of the special behaviour of why-pronouns (de Villiers 1991, 1996; Rizzi 2001; Thornton 2008 and many others).

---

1 English, German, Dutch, Swedish, French, Italian, Modern Greek, Russian, Ukrainian, Polish, Czech, Serbian, Bulgarian, Lithuanian, Finnish, Hungarian, Moksha, Western Mari, Turkish, Georgian, Basque.
Figure 1. The result of the PCA-analysis