Building a Corpus of Indefinite Uses
Annotated with Fine-grained Semantic Functions

Maria Aloni, Andreas van Cranenburgh, Raquel Fernández, Marta Sznajder

Institute for Logic, Language & Computation
University of Amsterdam
{M.D.Aloni|raquel.fernandez}@uva.nl
andreas.van.cranenburgh@huygens.knaw.nl,Marta.Sznajder@lrz.uni-muenchen.de

Abstract

Natural languages possess a wealth of indefinite forms that typically differ in distribution and interpretation. Although formal semanticists have strived to develop precise meaning representations for different indefinite functions, to date there has hardly been any corpus work on the topic. In this paper, we present the results of a small corpus study where English indefinite forms *any* and *some* were labelled with fine-grained semantic functions well-motivated by typological studies. We developed annotation guidelines that could be used by non-expert annotators and calculated inter-annotator agreement amongst several coders. The results show that the annotation task is hard, with agreement scores ranging from 52% to 62% depending on the number of functions considered, but also that each of the independent annotations is in accordance with theoretical predictions regarding the possible distributions of indefinite functions. The resulting annotated corpus is available upon request and can be accessed through a searchable online database.

**Keywords:** Corpus annotation, formal semantics, indefinites

1. Introduction

Natural languages possess a wealth of indefinite forms. English, for example, has at least four different indefinite determiners: *a, some, any, one*. Italian has many more including *uno, nessuno, (un) qualche, (uno) qualsiasi/qualunque, qualsivoglia*. These various forms seem to have a common logical/semantic core (their main function is to express indefinite reference), but typically differ in distribution and interpretation. For instance, there are contexts where English *some* and *any* can be interchanged (e.g. conditionals: *If you hear something/anything, call me*), while in other contexts using one or the other leads to different interpretations or to ungrammatical sentences (e.g. direct negations: *I didn't meet someone/anyone*; permissions: *You may kiss someone/anyone*; and episodic sentences: *I kissed someone/#anyone*). Italian determiner *qualsiasi* behaves like *any* in permissions and episodic sentences, but unlike *any*, it is ungrammatical in direct negations. German *irgendein* exemplifies yet another distribution/meaning pattern, resembling *any* in permissions, but being closer to *some* in episodic sentences.

Many theoretical questions arise from these observations. For instance, why is there so much cross-linguistic and language-internal variation in indefinite forms? What exactly is the common core of these various forms and what is specific to each of them? Why did some typological patterns emerge rather than others? As a starting point towards a principled answer to these questions, our group has conducted a number of synchronic and diachronic corpus studies of various indefinite forms cross-linguistically (Aguilar-Guevara et al., 2011). The main goal of this research is to understand and compare the meaning and distribution of these forms and to develop some hypotheses on their historical development. The point of departure for the identification of the relevant categories was the typological survey by Haspelmath (1997), who identified 9 main functions (context/meaning) for indefinite forms and organised them in an implicational semantic map (see Figure 1). Haspelmath proposes that an indefinite will always express a set of functions that are contiguous on the map (i.e. that form a connected sub-graph). A further prediction, relevant from a diachronic perspective, is that indefinites which acquire new functions will develop first those functions that are adjacent to the original function. One of the aims of our previous work (Aguilar-Guevara et al., 2011) was to test these hypotheses.

We extended the original map with a more detailed classification of negative and free choice uses of indefinites and developed a set of explicit logico-semantic tests organised in a binary decision tree, that would allow us to systematically assign particular functions on the map to instances of indefinites in context. Haspelmath's original map was extended as follows: the indirect negation function was split into an anti-morphic (AM) and an anti-additive (AA) function (to be able to distinguish between strong and weak negative polarity items, cf. (Zwarts, 1998)); and three new functions have been introduced contiguous to the free choice area, namely the generic function (GEN), the universal free choice (UFC) function, and the indiscriminative function (IND) (to be able, for example, to distinguish universal free-choice items like Italian *qualunque* from existential free-choice items like Italian *uno qualunque*). With these extensions in place, we then conducted several pilot synchronic and diachronic corpus studies where indefinite forms in Dutch, German, Spanish, Italian, and Czech where annotated by one annotator per language using the decision tree. These preliminary corpus studies confirmed Haspelmath's hypothesis of function contiguity.

The present paper makes the following contributions: (i) we extend previous work by developing annotation guidelines for the logico-semantic tests that can be used by non-expert annotators; (ii) we conduct a small corpus study on English indefinites *some* and *any* and report results of
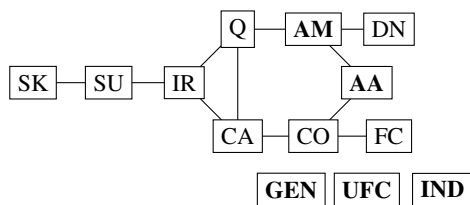
	Abbr		Example
SK	specific known		*Somebody* called. Guess who?
SU	specific unknown		I heard *something*, but I couldn't tell what it was.
IR	irrealis		You must try *somewhere* else.
Q	question		Did *anybody* tell you anything about it?
CA	conditional antecedent		If you see *anybody*, tell me immediately.
CO	comparative		John is taller than *anybody* else.
DN	direct negation		John didn't see *anybody*.
AM	anti-morphic		I don't think that *anybody* knows the answer.
AA	anti-additive		The bank avoided taking *any* decision.
FC	free choice		You may kiss *any* woman.
UFC	universal free choice		John kissed *any* woman with red hair.
GEN	generic		*Any* dog has four legs.
IND	indiscriminative		I don't want to sleep with just *anybody* anymore.

Figure 1: An extended version of Haspelmath's map (new functions in boldface) and a short description of the functions.

inter-annotator agreement; (iii) we make the resulting annotated corpus available to the research community through a searchable online database.

2. Data and Procedure

In this section we describe the methodology used to set up and carry out the annotation and the annotated dataset.

2.1. Guidelines

The aim of the methodology proposed by Aguilar-Guevara et al. (2011) was to come up with a systematic way of labelling uses of indefinites with one of the functions in the extended Haspelmath map in Figure 1. However, the logico-semantic tests that constitute the nodes of the decision tree were hardly usable by non-expert annotators (the tree is shown in Figure 3). For example, some of the tests referred to "the main operator of a sentence", without providing means to identify such an operator for any given sentence – a task that is easy only for trained linguists. One of such tests was the test to discriminate between universal and non-universal meaning (test **[c]** in Figure 3) . Originally this test was formulated as in (1), without any supplementary description for non-expert annotators.

(1) $\dots \mathbf{Op} (\dots \text{indefinite} \dots) \dots \Rightarrow \dots \forall x (\mathbf{Op} \dots x \dots) \dots$

In order to evaluate Aguilar-Guevara et al.'s methodology with multiple non-expert coders, we thus first set ourselves the task of defining annotation guidelines that would elucidate the tests. The guidelines include a description of each function in the map, the decision tree in Figure 3, and a description of the test to be applied at each non-terminal node of the tree, including an intuitive description of how to apply the test and some examples. For instance, in the guidelines the test for universal meaning was supplemented with a statement of the goal of the test and a detailed description of how to transform the target sentence into a form conforming to the scheme given in (1).[1] We illustrate the procedure of annotating one target indefinite in section 2.3. after introducing the annotation dataset.

2.2. Dataset

We constructed our annotation dataset by extracting 100 instances of indefinites from the British National Corpus through the BYU-BNC web interface (Davies, 2004). Of these, 80 items were instances of *any* and 20 of *some*. All items were independently annotated by 5 annotators, all of them graduate students at the ILLC. Only one of them was a native English speaker; the remaining 4 were proficient English speakers whose native language was Dutch (2 annotators), Russian (1) or Polish (1). The annotation scheme consisted of the functions in the extended Haspelmath map plus an additional label UN for "unclear", intended for cases where a test in the decision tree was inconclusive. Annotators were provided with the guidelines and received a few sessions of training where the guidelines were discussed. The annotation was done through an online interface that showed each indefinite to be annotated in context (100 tokens left and right, respectively).

2.3. An Example

Given the decision tree and the guidelines described in section 2.1., annotating a given item consists in applying a series of tests to a sentence containing the target indefinite, bearing in mind the context in which it appears. Consider one of the items in the annotation task (the target use of *any* is underlined and in italics):[2]

(2) **Item 80**: To avoid _*any*_ presumptions about the structure of the DNA, we replaced the bent DNA in the actual complex with the phosphate backbone for B-form DNA that was used to model the CAP/DNA complex.

Applying the first test, **[a]**, leads to the result S-, since it is impossible to continue the sentence containing the target indefinite with an episodic sentence starting with the pronoun *they* that would refer to "presumptions". The next test to be applied is therefore **[c]**, the test for universal meaning. The result of applying that test is positive, since from the given context a (rather artificial) scheme *For every presumption (about the structure etc.) x: to avoid x we replaced the bent*

---

[1]The guidelines are available at `http://staff.science.uva.nl/~maloni/Indefinites/corpus.html`.

[2]For simplicity, we have reduced the amount of context surrounding the target indefinite and given only the sentence that contains the target. As mentioned at the end of section 2.2., annotators had access to one hundred tokens left and right of the target.
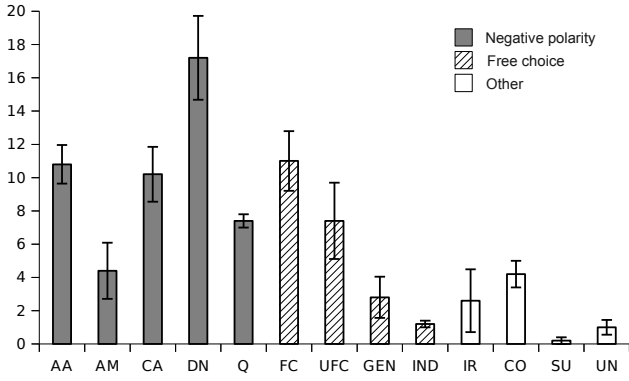
Figure 2: Average distribution of functions for *any* over 5 annotators. The error bars show the standard error.

Figure 5: Average distribution of functions for *some* over 5 annotators. The error bars show the standard error.

Q	85.0 %	FC	48.15 %
IND	83.33 %	IR	39.58 %
DN	73.26 %	UFC	37.5 %
CO	67.5 %	GEN	28.57 %
AA	66.67 %	AM	20.45 %
CA	64.0 %	SU	8.82 %
SK	61.90 %		

Table 1: Accuracy for individual functions, obtained by accumulating agreement scores among all possible pairs of annotators.

DNA etc. can be inferred. The result ∀+ leads to the test **[e]**, the test for anti-additivity, which comes out positive too. **[e]** requires the annotator to replace the phrase *any presumptions about the structure of the DNA* in (2) with a disjunction of two other NPs and check if a conjunction of corresponding sentences each containing only one of those NPs follows. For example in this case the sentences in question would be *To avoid problem1 or problem2, we replaced the bent DNA etc.* and *To avoid problem1, we replaced the bent DNA etc. and to avoid problem2, we replaced etc..* It is clear that the second sentence does indeed follow from the first. The next test to be applied is then **[g]**, the test for negative meaning. Since the target indefinite is not in the immediate scope of sentential negation, we construct a sentence *To avoid a presumption or no presumption, we replaced the bent DNA etc.* and observe that it is inconsistent. This leads to the result neg+ which leads us to the last test, **[h]**. The test for anti-multiplicativity comes out negative, since the sentence *To avoid problem1 and problem2, we replaced the bent DNA etc.* does not follow from *To avoid problem1, we replaced the bent DNA etc. or to avoid problem2, we replaced etc..* Therefore (2) can be classified as a case of anti-additive (**AA**) use of *any*.

3. Results and Analyses

Although the distributions of functions assigned to *some* and *any* differed across annotators, none of them violated Haspelmath's hypothesis of functional contiguity. Figures 2 and 5 show the average distribution of functions assigned by all the annotators to the 80 instances of *any* and the 20 instances of *some*, respectively. There were 6 instances classified as unclear (UN; see section 2.2.) by some annotator. For example, an annotator made use of this label to classify an instance of *any* that appeared within the idiomatic multi-word expression *at any rate*. As shown in figure 4 free choice uses of *any* were less frequent than its negative polarity uses.

We compared the annotations using *kappa* (Carletta, 1996; Artstein and Poesio, 2008). The overall *kappa* score was 0.52, with a standard deviation of 0.07 (*alpha* scores were equivalent up to the second decimal). Inter-annotator agreement was therefore moderate, which was not surprising given the fine-graininess of the annotation scheme.
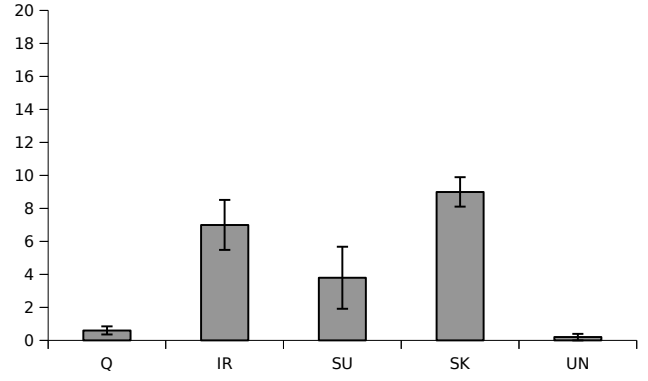
To get a sense of how much agreement there was for each individual category, we calculate the accuracy of category x with the following function:

$$f(x) = \frac{\sum_{a,b \in C \wedge a \neq b}^{a,b} \mid D_x^a \cap D_x^b \mid}{\sum_{a,b \in C \wedge a \neq b}^{a,b} \mid D_x^a \mid}$$

where C is the set of coders, and D_x^a gives the set of items annotated with category x by coder a. In this formula each coder takes a turn as being the gold (reference) set against which the other coders are evaluated. The resulting agreement scores are micro-averaged in a single accuracy score. In other words, the score represents the percentage of annotations which coincide when each annotator is compared to all others. The results are given in Table 1. The negative polarity labels score well, except for the new category AM. The free choice categories appear to be more difficult, except for IND for which there was one sentence which was unanimously identified.

An analysis of the pair-wise confusion matrices showed that for *some* there were frequent disagreements between SU, SK, and IR. The confusion between SU and SK is to be expected in English, since there isn't a grammaticalised distinction between these two functions in this language. A problematic example is shown in (3), which is compatible with a situation where the speaker knows Pamela's joke (SK) but also with one in which the speaker does not know it (SU):

(3) A laugh from the CARDINAL and the FOX at *some* witticism of PAMELA's.

We therefore also calculated agreement discounting the confusions between SU and SK by means of the following
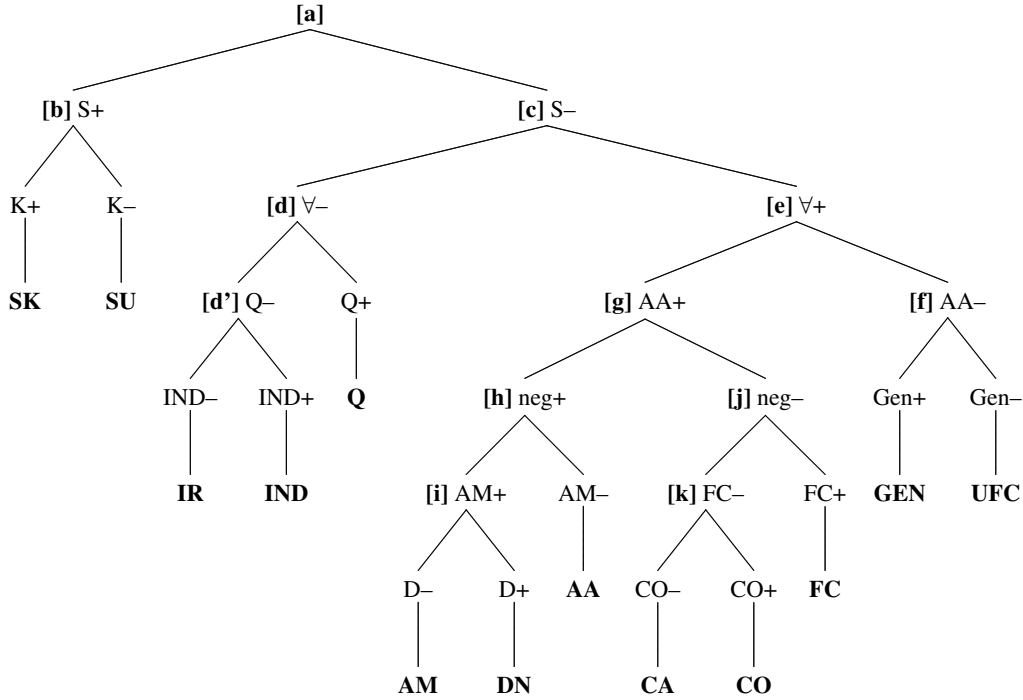
Figure 3: Decision tree proposed by Aguilar-Guevara et al. (2011).

(a) *specificity* [S+/–]: does the given indefinite refer to a specific individual in the actual world, i.e. is it possible to continue the given sentence with an episodic sentence containing a personal pronoun which is anaphoric to the given indefinite?

(b) *known* [K+/–]: does the speaker know who the referent of the indefinite is, e.g. is it possible to continue the given sentence with "Guess who/what?"?

(c) *universal meaning* [∀+/–]: does the given sentence state something about a single individual or can it be applied to all individuals in the domain specified by the sentence? I.e. does the given sentence $\mathbf{Op}(\ldots \text{indefinite} \ldots)$ imply the universal sentence $\forall x\, \mathbf{Op}(\ldots x \ldots)$?

(d) *polar question* [Q+/–]: is the indefinite embedded under a polar ("yes/no") question operator?

(d') *indiscriminacy* [IND+/–]: does the existential non-specific reading have a free choice (arbitrary) flavor?

(e) *anti-additivity* [AA+/–]: does the relevant sentential operator $\mathbf{Op}$ have the property of anti-additivity? I.e. does it hold that $\mathbf{Op}(a \vee b) \Rightarrow \mathbf{Op}(a) \wedge \mathbf{Op}(b)$?

(f) *genericity* [Gen+/–]: does the sentence express some generalisation about the *kind* of individuals denoted by the indefinite?

(g) *negative meaning* [Neg+/–]: does the relevant operator $\mathbf{Op}$ have a negative meaning? I.e. is $\mathbf{Op}(a \vee \neg a)$ inconsistent?

(h) *anti-multiplicativity* [AM+/–]: does the relevant operator $\mathbf{Op}$ have the property of anti-multiplicativity? I.e. does it hold that $\mathbf{Op}(a) \vee \mathbf{Op}(b) \Rightarrow \mathbf{Op}(a \wedge b)$?

(i) *clausal negation* [D+]: is the negation in the same clause as the indefinite?

(j) *free choice* [FC+/–]: is there a free choice between certain (maybe opposite) alternatives? I.e. is $\mathbf{Op}(a \vee \neg a)$ is informative, where $\mathbf{Op}$ is the relevant operator?

(k) *comparative construction* [CO+]: does the indefinite occur as part of a comparative construction?

Figure 4: Outline of the tests referenced in Figure 3.

distance function, which treats them as a single label:

$$(4) \quad d(a,b) = \begin{cases} 0 & \text{if } a = b \\ 0 & \text{if } a,b \in \{\text{SK}, \text{SU}\} \\ 1 & \text{otherwise} \end{cases}$$

When these two categories are collapsed into one in this manner, we obtain a *kappa* score of 0.56 (with 0.07 standard deviation).

As for IR, ambiguity seems to have played an important role. Some uses of *some* are often ambiguous between IR and SU/SK. For instance, in (5) two annotators chose label SK while the other three chose IR. In this case, the IR reading seems the most natural one, but there is also a possible reading according to which the speaker has in mind a specific period of time.

(5) I will allow you to lie half an hour after me then you'll have <u>*some*</u> time you may call your own.

Ambiguity of this sort was difficult to detect by the annotators, which sometimes led to situations where disagreements arose because annotators had focused on different readings. Another case in point is the instance of *any* shown in (6).

(6) For if God could have made the world work in <u>*any*</u> number of ways, would it not always be presumptuous to pretend that one had actually pinned Him down?

Sentence (6) is ambiguous between a free choice (universal) interpretation, which can be paraphrased as : "If God could have made the world work in every number of ways, . . . "; and a negative polarity (existential) interpretation: "If

God could have made the world work in some number of ways, . . .". In the given context the free choice interpretation seems the most natural one. When applying our decision procedure to this example, at node **(c)** (the test for universal reading) we have to decide on what operator counts as the relevant **Op**. We have two plausible candidates here: the conditional construction or the modal *could*. In the first case (corresponding to the negative polarity reading) our terminal node will be CA. In the second case (corresponding to the free choice reading) our terminal node will be FC. Three of our annotators chose for a free choice interpretation (FC or UFC), one for the negative polarity interpretation (CA), and one labeled the sentence as ambiguous between FC and CA.

Regarding *any*, most of the disagreements concerned the fine-grained functions that had been added to the original map proposed by Haspelmath: in the extended version of the map the classification of negative uses had been made more precise by adding AM and AA in place of Haspelmath's IN (indirect negation), while FC had been complemented by UFC, GEN, and IND. If we collapse these two groups of functions and thus consider the original Haspelmath's map, inter-annotator agreement increases substantially, with a *kappa* score of 0.62 (and a standard deviation of 0.05). The tests developed to distinguish AA from AM, and FC from UFC/GEN rely on intuitions about entailments of (embedded) disjunctive sentences (see tests **(e)** and **(h)** in Figure 4). Reasoning tasks involving disjunction are known to be cognitively hard. Furthermore there is a lot of cross-linguistic variation with respect to the possibility of embedding disjunction. This might explain why these newly introduced distinctions led to disagreements between our annotators, who had a variety of native languages.

4. The Corpus

The work presented in this paper is part of our ongoing effort to create a cross-linguistic corpus of indefinite uses annotated with fine-grained functions as identified by formal semanticists. We expect the corpus to be a valuable resource for conducting synchronic, diachronic, and typological studies of the different form/function mappings exhibited by lexical items used to express indefinite reference. For now, we make available the English corpus described in this paper together with the multi-coder annotation. Eventually the corpus will also include annotated uses of German *irgend*-indefinites (synchronic and diachronic), Spanish *cualquier*-indefinites (synchronic and diachronic), Dutch *wh dan ook*-indefinites (synchronic and diachronic), Italian *unque*-indefinites (synchronic) and Czech *koli*-indefinites (synchronic).

The corpus is accessible through an online interface that allows users to browse the corpus restricting several parameters, including document genre; to search for items annotated with particular functions (by one or more annotators); and to download the dataset and/or the annotations. A beta version of the online interface is available from: `http://staff.science.uva.nl/˜maloni/Indefinites/corpus.html`.

5. Conclusions

Although indefinites have been extensively studied within the field of formal semantics, to date there has hardly been any corpus work on the topic. The research we have presented in this paper can be seen as a first step towards bridging this gap. We have reported the results of a small corpus study on the semantic functions of English indefinites *some* and *any*. One of the main challenges involved in the study has been the development of guidelines for the annotation of fine-grained semantic functions that could be used by non-expert coders. We have calculated inter-annotator agreement amongst a set of 5 annotators obtaining moderate *kappa* scores that range between 52% and 65% depending on the number of functions considered. Although the distributions of functions assigned to *some* and *any* differed across annotators, none of them violated theoretical predictions regarding the possible distributions of indefinite functions (Haspelmath's hypothesis of functional contiguity). As mentioned in the previous section, the corpus and the annotations are readily available and will soon be extended with studies on indefinite uses in several other languages.

7. References

A. Aguilar-Guevara, M. Aloni, A. Port, R. Šimík, M. de Vos, and H. Zeijlstra. 2011. Semantics and pragmatics of indefinites: methodology for a synchronic and diachronic corpus study. In *Proceedings of the DGfS Workshop "Beyond Semantics: corpus-based investigations of pragmatic and discourse phenomena"*. BLA.

R. Artstein and M. Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

J. Carletta. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2):249–254.

M. Davies. 2004. BYU-BNC: The British National Corpus. Available at `http://corpus.byu.edu/bnc/`.

M. Haspelmath. 1997. *Indefinite Pronouns*. OUP.

F. Zwarts. 1998. Three types of polarity. *Plurality and quantification*, 69:177–238.