

	Cohen's $\kappa$
Before	0.69
During	0.59
After	0.77
All	0.70

Table 2: Inter-annotator agreement (Cohen's kappa). Values over 0.6 are considered *substantial*, over 0.8 are considered *perfect* (Artstein and Poesio, 2008).

over 75% of articles have at least 19 unique potential possessors, 5 years and 46 pairs; and over 50% of articles have at least 28 unique potential possessors, 8 years and 86 pairs. In other words, our corpus takes into account many potential possessors and years for the vast majority of articles.

### 3.3 Validating Possessors and Years

After (potential possessor, year) pairs were generated, they were validated manually. To do so, we asked the following questions to annotators:

- Did a possession relation exist between the potential possessor and the possessee at any point of time *before* year?
- Did a possession relation exist between the potential possessor and the possessee at any point of time *during* year?
- Did a possession relation exist between the potential possessor and the possessee at any point of time *after* year?

In all questions, *possessee* refers to the artwork the Wikipedia article is about. Annotators had to choose from two answers: *yes* or *no*, where *no* indicates all cases in which there is not enough information to determine that a possession relation exists at any point of time before / during / after year. In other words, *no* does not mean that the potential possessor did not possess the possessee, and it may mean that there is no information about whether a possession relations exists.

The annotation interface showed the title of the article and the section to which the potential possessor and year belong to (section title + text). Annotators were instructed to first read the section and then answer all questions. Thus, annotators reveal possession information involving possessors and years that are potentially far away (different clauses, sentences, etc.). Recall that all potential possessors and years within a section are paired, thus we allow to cross sentence boundaries.

**Annotation Quality.** Annotations were done in-house by two graduate students. Both of them

annotated 25% of the articles individually. Table 2 shows inter-annotator agreements (Cohen's kappa) for each question. Overall, inter-annotator agreement is 0.70 (values between 0.60 and 0.80 are considered substantial (Artstein and Poesio, 2008)). Agreements are higher for *Before* and *After* than *During* (0.69 and 0.77 vs. 0.59). The remaining articles were annotated once.

**Annotation Examples.** Figure 3 shows the annotations for one paragraph of the Wikipedia article about *Girl with a Pearl Earring* (more specifically, from the section titled *Ownership and display*). The figure shows the annotations on top of a screenshot of the article for clarity purposes, but the annotation interface only showed one section at a time along with all the generated pairs (Section 3.3, equivalent to pre-drawing edges).

Five potential possessors and two years were selected, thus ten (potential possessor, year) pairs were generated. The annotations reveal the intuitive possession information contained within the paragraph. First, *Victor de Stuers* was an advisor to *Arnoldus Andries des Tombe*, so there is no evidence that he was a possessor at any point of time (missing label edges). Second, *Vermeer* is the artist who made *Girl with a Pearl Earring*, so there are possession relations before 1881 and 1902. Third, *Arnoldus* purchased the piece in *The Hague* in 1881, and in 1902 it was donated to *Mauritshuis*. So *Arnoldus* was a possessor in 1881 and after 1881 (until 1902), *The Hague* in 1881 (recall that non-humans can be possessors, spatial proximity is also considered possession, Section 2.1), and *Mauritshuis* during and after 1902. We discuss the limitation of the annotation approach in Section 3.5.

### 3.4 Annotation Analysis.

Counts of *yes* labels for the three questions (before, during and after) are rather low (17%, 9% and 19%, Figure 4). This is not surprising, as any PERSON, ORG, LOC and GPE named entity is considered as potential possessors. We note, however, that we annotated a possession relation (*yes* label) in 35% of (potential possessor, year) pairs generated (either before, during or after year).

Figure 5 depicts the distribution of labels per article for (potential possessor, year) pairs generated from the same and different sentences. It is worth noting a couple of interesting patterns. First, the annotations contain many more possessions because we pair potential possessors and years that