Temporally-oriented possession: A corpus for tracking possession over time

Temporally-oriented possession: A corpus for tracking possession over time

Dhivya Chinnappa, Alexis Palmer and **Eduardo Blanco**
University of North Texas
dhivyainfantchinnappa@my.unt.edu, alexis.palmer@unt.edu, eduardo.blanco@unt.edu

1 Temporally-oriented possession

From a linguistic perspective, the term *possession* refers to a set of semantic relations between two entities, the *possessor* and the *possessee* (Stassen, 2009). A wide range of different asymmetric relationships fall under the heading of possession, including kinship, proximity, part-whole relations, experience of abstract concepts, and physical possession, both permanent and temporary. The most typical notion of possession involves ownership or control of the possessee by the possessor, as in phrases like "my piano," "the lion's beautiful tail," or "this friend of mine." The linguistic literature makes a conceptual distinction between *alienable possession*, in which possessees can be separated from their possessors, and *inalienable possession*, in which such separation is not possible (Aikhenvald and Dixon, 2012; Heine, 1997, among others). Unlike inalienable possessions, which are permanent, alienable possessions are temporary and, therefore, capable of changing hands. We are interested in tracking change of possessions.

Previous work on automatic extraction of possession has mostly focused on particular syntactic constructions. Tratz and Hovy (2013) investigate various semantic relations realized by English possessive constructions, and both Nakov and Hearst (2013) and Tratz and Hovy (2010) consider possession expressed by noun compounds. Badulescu and Moldovan (2009) extract possession as one of the many semantic relations expressed by English genitives. Blodgett and Schneider (2018) present a corpus of web reviews annotating genitives with adpositional supersenses, finding that this inventory works for canonical possessives. We consider *all* expressions of possession, whether phrasal, clausal, or sentential, or even inter-sentential. The non-restrictive approach presented here is similar to that of Banea et al. (2016), who annotate possessions of bloggers at the time of utterance. In our previous work (Chinnappa and Blanco, 2018), we first extract possessions from a sentence using a deterministic procedure, and then identify the types and temporal anchors of possession.

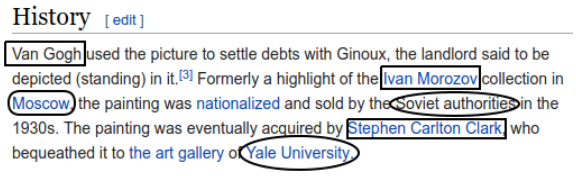
History [edit]
Van Gogh used the picture to settle debts with Ginoux, the landlord said to be depicted (standing) in it.[3] Formerly a highlight of the Ivan Morozov collection in Moscow, the painting was nationalized and sold by the Soviet authorities in the 1930s. The painting was eventually acquired by Stephen Carlton Clark, who bequeathed it to the art gallery of Yale University.

Figure 1: Excerpt from the Wikipedia article *The Night Cafe*. The possessors are highlighted.

We present a new corpus of Wikipedia articles annotated with *temporally-oriented possession* or tracking concrete objects as they change hands over time. Our corpus consists of 90 Wikipedia articles (Figure 1). All artifacts and possessors are concrete entities, with possessors limited to people, organizations, and locations.

2 Data: Articles about famous artifacts

We collected a corpus of English Wikipedia articles about historical artifacts that could possibly change hands over time, being held by different possessors in different years. The article topics included paintings, diamonds, relics, sculptures, and archeological findings.

Next, the set of articles was filtered to retain only articles that: a) focus their discussion on a single artifact; and b) contain at least three possessors for the artifact. These filtering criteria are motivated by our end goal of automatically extracting possession timelines from the texts. The resulting corpus consists of 90 articles, with each article focusing on a single target artifact. Table 1 shows basic statistics for the corpus. The data and annotations are freely available.[1]

[1] Available at dhivyachinnappa.com.

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 302-305.
New York City, New York, January 3-6, 2019

Total # of Wikipedia articles	90
Total # of mentioned possessors	799
Total # of unique possessors	735
Avg # of words per article	2315
Avg # of sections per article	6.66
Avg # of possessors per article	8.87
Avg # of unique possessors per article	8.17

Table 1: Corpus statistics

All possessors	
Total for all articles	799
NE type: Per / Org / Loc	281 / 318 / 200
Possession Certainty: Cert / Uncert	774 / 25
Temporal Anchor: Known / Unknown	**660** / 139
Duration: Before / During / After	7 / 647 / 6
Duration Certainty: Cert / Uncert	608 / 52

Table 2: Statistics for all marked possessors

3 Annotating possession

The annotation scheme was designed primarily to capture all temporal information relevant to changes of possession over time. Thus, in addition to identifying artifact-possessor relations (Section 3.2), we identify a temporal anchor for each relation and the duration of the possession with respect to the temporal anchor (Section 3.5). The set of possession relations is then ordered into a timeline (Section 3.6). For both possession relations and possession duration, we annotate whether these features are certain or not, given the available textual evidence (Section 3.4).

Annotators were provided with HTML pages of the 90 selected Wikipedia articles as downloaded on 12th June 2017. Annotation was done using the Wired-Marker[2] Firefox extension to annotate the HTML pages. First, all possessors of the target artifacts (Section 3.2) were highlighted, using different-colored markers (provided by Wired-Marker) for different named entity types (Section 3.3). All other annotation features (Sections 3.4, 3.5, 3.6) were added to the highlighted text using Wired-Marker's notes function.

3.1 General instructions to annotators

The annotators were instructed to read the entire document to decide on the possessors and the order of possession. Unless a possessor possessed the artifact at different points in time, only one mention of each possessor is annotated.

3.2 Possessors and artifacts

We focus on a single artifact, namely the topic of the article. For that artifact, we identify all possessors *of that artifact* mentioned over the course of the article. The corpus consists of all artifact-possessor pairs identified from the selected Wikipedia articles. We extract 799 pairs in all, with 735 unique artifact-possessor pairs.

[2]http://www.wired-marker.org/en

3.3 The role of named entities

The possessors identified each fall into one of three named entity (NE) categories: Person, Organization (e.g. museums or universities), or Location (e.g. particular cities, states, or countries). The NE type of each possessor is labeled manually, with the resulting distribution shown in Table 2. Organizations are the most frequent possessors, followed by People and then Locations. Although the possessors fall neatly into traditional NE categories, many of them are not in fact recognized by standard NE taggers. These include cases like example (1):[3]

(1) On the morning of March 18, 1990, **thieves** disguised as police officers broke into the museum and stole *The Storm on the Sea of Galilee* and 12 other works. *[The Storm on the Sea of Galilee]*

The thieves who stole the painting, and presumably possessed it for at least some time thereafter, are unnamed. English NER systems also struggle to recognize possessors such as "artist's daughter" or names in other languages. This means that NER alone is not sufficient to identify possessors, even in this specific context where all possessees are artifacts and likely to be owned by NEs.

3.4 Certainty of possession

For each artifact-possessor pair, annotators are asked to assess the certainty of the possession relation. We are interested in the notion of certainty as it relates to textual evidence: if the text *of the entire article* strongly supports the relation, the instance should be marked as Certain (C). If not, it should be marked as Uncertain (UC). Nearly all relations are marked as certain (see Table 2).

Example (2) illustrates a case of uncertainty — the phrase "generally accepted" indicates some degree of uncertainty on the part of the author.

(2) It was completed after Giorgione's death in 1510, [...] generally accepted to have been completed by **Titian**. [*The Sleeping Venus*]

[3]Possessors (and sometimes temporal anchors) appear in boldface, and the article name in square brackets.

NE	Possessor	Poss.Cert	Order	Anchor	Duration	Dur.Cert
PER	Vincent van Gogh	C	1	1888	During	C
PER	Ivan Morozov	C	2	Unknown	-	-
PER	Stephen Carlton Clark	C	4	Unknown	-	-
ORG	Soviet authorities	C	3	1930	Before	C
ORG	Yale University	C	5	Unknown-Now	During	C
LOC	Moscow	C	2	Unknown	-	
LOC	New Haven, CT	C	5	Unknown-Now	During	C

Table 3: Complete annotation for the Wikipedia article on Van Gogh's *The Night Café*.

3.5 Temporal anchor and duration

The core of our approach is to annotate temporal features of the extracted possession relations.

Temporal anchor. The first time-related annotation decision is to determine whether, according to the text, there is a temporal anchor for the given possession relation. For cases when a possessor has held an artifact for more than one time period, different temporal anchors may be associated with the same artifact-possessor pair, as in example (3) below. This painting was in the custody of its owner prior to the 1873 Exhibition, and then again for a period between the end of that exhibition and the painting's 1878 journey out of Russia.

(3) Despite its progressive implications, Barge Haulers was bought by the **Tsar's second son**. It was lent for exhibition at the 1873 **International Exhibition** in **Vienna**, where it won a bronze medal. It was exhibited **outside Russia** again in 1878... *[Barge Haulers on the Volga*]

If a temporal anchor cannot be identified, the other temporal features are not relevant. Looking again at Figure 1, only one of the five possessors has an identifiable temporal anchor: **Soviet authorities** and **1930**. The ordering of possessors is clear, but only one of three possessions has a specific time. In the corpus, 660 of 799 possessors are associated with a temporal anchor.

Duration of possession. Our temporal anchors similar in nature to the TLINK annotations of the TimeBank Corpus (Pustejovsky et al., 2003), but we restrict the granularity to the level of year. An anchor could denote a single year, a range of years, or some historical event. Particular days or months are ignored.

In the ideal case, the temporal anchor covers the entire period of possession (e.g. "1983-1987"), but more often the text mentions a date or historical event (e.g., World War II) which may or may not lie within the duration of possession. To build possession timelines, we need to know how the temporal anchor relates to the period of possession. Thus we annotate three different categories of duration. BEFORE indicates that possession occurred prior to the anchor, while AFTER indicates that possession occurred later than the temporal anchor. DURING indicates that the period of possession includes the temporal anchor.

(4) BEFORE: At some undetermined point before **1516** it came into the possession of **Don Diego de Guevara** ... [*The Arnolfini Portrait*]

(5) DURING: In **1599** a German visitor saw it in the **Alcazar Palace** in **Madrid**.[*The Arnolfini Portrait*]

(6) AFTER: In **1530** the painting was inherited by Margaret's niece **Mary** of Hungary, who in 1556 went to live in Spain. [*The Arnolfini Portrait*]

Note that the duration annotations reflect only the knowledge contained in text; they do not provide complete information about changes in possession. The temporal anchor in example (4) provides the latest possible date at which possession of the portrait transferred to Don Diego de Guevara. Example (5) conveys that the temporal anchor 1599 occurs sometime during the Alcazar's possession of the portrait. We do not know where in the period of possession the date falls; it could be a beginning or end date. The temporal anchor in example (6) marks the change of possession.

3.6 Possession ordering and timeline

The final annotation task is to order the possessors according to when each had control of the artifact in question, building up a **possession timeline**. Each possessor is given a serial number, depending on the order in which the artifact was possessed. Usually the artist who created or found the artifact (if known) is assigned the serial number 1. The next possessor gets the serial number 2, and so forth. An example timeline can be seen in Table 3.

Annotation relies on the complete textual context, often allowing the annotator to determine or-

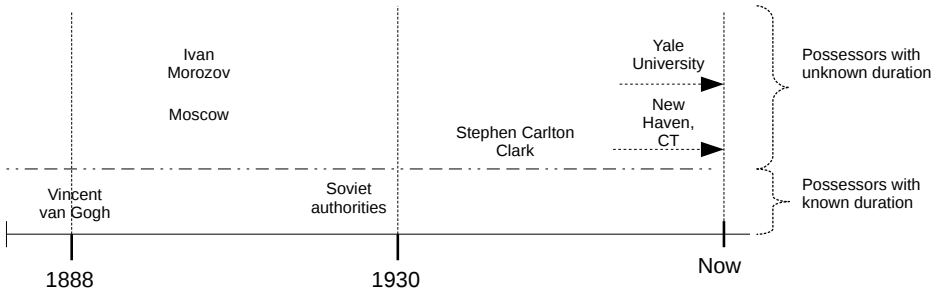

Figure 2: Possessors placed on a timeline using information from annotations from Table 3.

dering of possession events even if explicit temporal anchors are not stated in the text. For example, in Figure 1, we can easily infer that the possessor (**Yale University**) to whom the painting was bequeathed appears in the timeline later than the possessor who did the bequeathing.

When two possessors held the artifact simultaneously (e.g. **Yale University** and **New Haven, Connecticut**), both appear at the same position in the timeline. A possessor can receive multiple serial numbers when there are multiple relevant periods of possession (as in example (3)), and repeated mentions of possessors tied to the same period of possession are marked as (e.g.) 1.1, 1.2, 1.3.

4 Discussion

For validation, twelve randomly-selected articles are labeled by a second annotator. We treat Annotator A's labels as a pseudo-gold standard and measure precision and recall of Annotator B's labels as compared to Annotator A. For identification of artifact-possessor pairs, precision is 0.97, and recall is 0.69. Inter-annotator agreements for temporal and certainty features is calculated only for the set of artifact-possessor pairs identified by both annotators. For certainty (both possession and duration), Cohen's κ is very high (0.92). Agreement is more moderate for the temporal features; Cohen's κ of 0.77 for temporal anchor and 0.76 for duration of possession. For the order of possession, we generate a list of ordered pairs of possessors for both annotators and then compare. Precision between the two lists of pairs is 0.93, and recall is 0.90.

The goal of this corpus is to enable further research: a) to better understand the nature of changes in possession over time; b) to analyze how such possession changes are realized in text; and c) to lay a foundation for automatic extraction of possession timelines. Extracting temporally anchored possessions may be useful in analyzing and understanding the history of the artifacts, as well as for enriching more general event timelines.

References

A.Y. Aikhenvald and R.M.W. Dixon. 2012. *Possession and Ownership: A Cross-Linguistic Typology*. Explorations in Linguistic Typology.

Adriana Badulescu and Dan Moldovan. 2009. A semantic scattering model for the automatic interpretation of english genitives. *NLE* .

Carmen Banea, Xi Chen, and Rada Mihalcea. 2016. Building a dataset for possessions identification in text. In *LREC*.

Austin Blodgett and Nathan Schneider. 2018. Semantic Supersenses for English Possessives. In *LREC*.

Dhivya Chinnappa and Eduardo Blanco. 2018. Mining possessions: Existence, type and temporal anchors. In *NAACL: HLT*.

B. Heine. 1997. *Possession: Cognitive Sources, Forces, and Grammaticalization*. Cambridge Studies in Linguistics. Cambridge University Press.

Preslav I. Nakov and Marti A. Hearst. 2013. Semantic interpretation of noun compounds using verbal and other paraphrases. *ACM TASLP* .

J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003. The TIMEBANK corpus. In *Corpus Linguistics 2003*.

L. Stassen. 2009. *Predicative Possession*. Oxford Studies in Typology and Linguistic Theory.

Stephen Tratz and Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. In *ACL*.

Stephen Tratz and Eduard H. Hovy. 2013. Automatic interpretation of the english possessive. In *ACL*.