

University of Groningen

A semantically annotated corpus of tombstone inscriptions

Bos, Johan

Published in:
International Journal of Digital Humanities

DOI:
[10.1007/s42803-021-00039-y](https://doi.org/10.1007/s42803-021-00039-y)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Bos, J. (2022). A semantically annotated corpus of tombstone inscriptions. *International Journal of Digital Humanities*, 3(1-3), 1-33. <https://doi.org/10.1007/s42803-021-00039-y>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



A semantically annotated corpus of tombstone inscriptions

Johan Bos¹

Received: 15 June 2020 / Accepted: 11 October 2021 / Published online: 25 October 2021
© The Author(s) 2021

Abstract

The digital preservation of funerary material is of interest to many different scientific disciplines. Textual information found on tombstones often goes far beyond the expected (name of the deceased, dates of birth and death), and may include information about commemorators, family roles, occupations, references to biblical or other texts, places of birth and death, cause of death, epitaphs and poems. Gravestones are multi-modal media, and besides text are often decorated with artistic symbols. To capture this information in a systematic way and make it available on a large scale for research purposes, a meaning representation based on linking entities by relations has been designed that will extend search capabilities beyond simple string matches. Concepts are represented as WordNet synsets, and a vocabulary of 32 relations make connections between concepts. This formalisation has been developed and evaluated based on a dataset of more than 1,000 Dutch tombstones.

Keywords Data collection · Funerary artefacts · Data formalisation

1 Introduction

The digital preservation and dissemination of funerary material is of interest and great value to researchers of many different disciplines: anthropologists, archaeologists, art historians, cataloguers, cultural historians, genealogists, gender theorists, library scientists, linguists, philologists, sociologists, theologians and perhaps more (Saller & Shaw, 1984; Eckert, 1993; Carmack, 2002; Streiter et al., 2007; Veit & Nonestied, 2008; Long, 2016). Inscriptions on tombstones are a rich source of information for all of these areas of scientific study. However, the objects of study, the tombstones, face a constant threat of disappearance, caused by the clearance of

✉ Johan Bos
johan.bos@rug.nl

¹ Center for Language and Cognition, University of Groningen, Groningen, Netherlands

graveyards, obscurity caused by weathering (Matthias, 1967; Streiter et al., 2007), urban development, landslides, social or religious changes (Streiter et al., 2010a) or even political incentives (Streiter & Goudin, 2011).

Fortunately, several initiatives have been launched to ensure the digital preservation of tombstones and cemeteries (Streiter et al., 2010a; Toscano, 2019; Tobin et al., 2019). While many of these efforts are extremely valuable, they only target the first step of digitalisation. We believe that more can and should be done. In this article, we go a step further and propose a formalism that offers a semantic perspective on funerary material by capturing the meaning of tombstone inscriptions. The ultimate aim is to provide researchers of the above-mentioned disciplines with search capabilities that go beyond simple textual string matches.

This form of semantic search allows researchers to search for specific characteristics of their objects of study. For example, they can search for tombstones with pictures of animals or, more specifically, birds or butterflies; stones with names composed of patronymic suffixes; geographical locations of gravestones that refer to Psalm 103; stones where the deceased were married and where the wife is mentioned before the husband; Frisian headstones of teachers or priests; graves from a certain period with weeping trees; and so on. To enable this kind of semantic search for digital humanities research, formal descriptions are required that are ideally compatible with the Resource Description Framework (RDF) data model and other ongoing efforts on the semantic web (Antoniou et al., 2012).

The objectives of this paper are in line with the foundational work of Streiter et al. (2007), who propose an XML-based annotation scheme for graveyards and tombstones, focusing on the physical properties of tombstones. In our work, we aim to add structure to the representation of tombstone *inscriptions*. We extend the work of Streiter et al. (2007) by associating a tombstone inscription with a meaning representation that provides a model of the entities mentioned on the stone (people, dates, locations, symbols, religious references) and the relations between them (temporal relations, family relations, occupational relations), maintaining the difference between meaning and reference that Streiter et al. (2007) and Streiter et al. (2008) emphasise to be important to include in a corpus of tombstone annotations.

As the saying goes, there is no better data than more data, and large datasets are more easily produced with the help of computational algorithms to save time and money. In this article, we not only propose a formalism to capture the semantic dependencies of tombstone inscriptions, but also present a corpus of 1,100 semantically annotated images of funerary artefacts. This dataset is an important stepping stone for acquiring larger datasets of annotated data, and we envisage that machine learning could play a pivotal role in this (Tobin et al., 2019). As far as we are aware, this is the second large-scale corpus of tombstones and the first one of European tombstones, since Streiter et al. (2008) presented their corpus for Taiwanese tombstones.

This brings us to the research questions that we aim to answer in this article: What information is provided on tombstones in our study area (the north of the Netherlands)? What expressive power is required for a semantic formalism to represent this multi-modal information? Which vocabulary is required to describe this data in a formal meaning representation? And, finally, what challenges will be faced by



Fig. 1 Churchyard on Vlieland on 20 February 2020 (left) and the Zuiderbegraafplaats graveyard in the city of Groningen on 1 November 2019 (right)

computational approaches that have as their objective to learn these semantic annotations automatically?

To answer these questions, we first describe the process of data collection and give an informal characterisation of the types of information found on the tombstones in our dataset (Section 2). Next, in Section 3, we motivate the choice of semantic formalism. We proceed with describing how entities are conceptually instantiated and how semantic relations are formally captured (Section 4). Finally, in Section 5, we discuss the challenges that automatic tombstone reading approaches will have to deal with.

2 Collecting and interpreting tombstone data

2.1 Sources

There are several online databases with images of tombstones that would be appropriate for our purposes. However, as we are mainly interested in the information engraved on the stone, many of these pictures are not useful for our study because, too often, the texts are largely illegible. A second reason for collecting a new, raw dataset is to avoid the problems caused by intellectual property and ownership issues. This ensures that anyone can use our dataset for research purposes, as long as privacy issues of the deceased and their commemorators are taken into account.¹

Given these considerations, tombstone inscription data was collected by visiting 40 publicly accessible cemeteries in the three northern provinces of the Netherlands (Groningen, Drenthe and Friesland, see Fig. 1). Appendix A lists all of the visited graveyards and churchyards. The cemeteries contain stones of the 18th, 19th, 20th and 21st centuries. The major religion in this part of the Netherlands is Protestantism, and

¹The General Data Protection Regulation (GDPR) in the EU only applies to information relating to identifiable living individuals. Information found on publicly available tombstones related to deceased individuals is therefore not subject to the GDPR. However, to respect the commemorators of the deceased, we ask users of the dataset not to display information in publications (including websites) of persons who have recently passed away.

this is reflected in the information found on the stones. Only one of the visited cemeteries was exclusively dedicated to the Roman Catholic Church. Although there are several Jewish and Muslim cemeteries in this area of the Netherlands, they were left out of the current study because of the expertise required to transcribe and interpret scripts written in Arabic or Hebrew.

2.2 Digitalisation

Using a Samsung A50 smartphone, high-resolution pictures of tombstones were taken from a close distance, without the use of a flash, light or tripod. Images were stored together with geolocation (using GPS, Global Positioning System) and time and date information. Files were saved in JPEG format, their orientation adjusted when needed, and irrelevant parts of the background were manually cropped from the image (Fig. 2). Each file was named with a unique identifier. In total, more than 1,000 images were collected in this way. Most of these are images of headstones; some of them capture crosses or other burial artefacts used for commemoration.

Photographs were not taken if this interfered with mourners visiting the graveyard. If memorial objects or vegetation obstructed part of the view of the inscriptions, the stones were disregarded instead of temporally removing these objects for a full view of the tombstone. On rare occasions (fewer than 10 times), leaves, branches or small pieces of dirt were removed from stones to obtain a better photograph.



Fig. 2 Original tombstone image (left) and cropped version (right)

Not all of the tombs were photographed during a visit to a cemetery. Some stones, especially those crafted from marble, presented challenges to the photographer as the reflective surface of marble tombstones made photographing them difficult. Damaged stones or stones with unreadable inscriptions were also disregarded. For the sake of simplicity, funerary objects with information that required more than one image for full capture (e.g., obelisks or headstones with information on the front and back) were either ignored or selectively approached by taking only a single picture of the most important part. Tombs were selected with variety in mind, to build a diverse dataset with respect to language, symbols, usages of names, mentions of occupations, causes of death and family relations.

2.3 Information on stones: a concise, informal characterisation

Tombstones can generally be divided into family gravestones and headstones for individuals. Family gravestones are relatively rare and usually contain no more information than the name of the family. Individual stones contain information about one or more deceased person(s), usually with the dates of birth and death, and in rarer occasions, the date of death and the deceased's age in years and months (fewer than 3% of the stones in our dataset). Other persons mentioned are the commemorators, commonly in connection with a characterisation of the social relation with the deceased. Other types of information are the person's job or organisation they worked for, references to biblical texts, epitaphs, poems, cause of death and the name of the designer of the tombstone.

There is an abundance of linguistic variation for all of these types of information, with use of anaphoric and deictic pronouns, graphical coordination and coreference. Numeral expressions are used, but explicit universal quantification, disjunction or negation was not encountered. In just the three provinces of the northern Netherlands, seven languages were found: Dutch (obviously), plus Frisian (5x), Gronings (3x), English (5x), French (3x), Malay (2x) and Hebrew (1x).

The textual information is often decorated with symbols (more than half of the tombstones in our dataset) and sometimes portraits of the deceased (less than 1% in our dataset). The symbols are variations on common burial themes (weeping willows, broken flowers, winged sandglasses, anchors, axes, books, scythes, skulls, bones), animals (birds, snakes, butterflies, caterpillars, moths), religious or cult icons (Star of David, pentagram, swastika) or more contemporary artefacts (musical instruments, boats, cars, bicycles).

3 Information on stones: a formal characterisation

3.1 Choice of formalism

The quest of deciding on a meaning representation for a particular modelling task boils down to finding a healthy balance between expressive power and inferential tractability. As we have seen in our informal description of our domain, tombstone inscriptions require modelling entities of various kinds (persons, locations, date

expressions) and the relations between these (family relations, occupations, references). Since negation and disjunction do not seem to play an important role, a simple graph-like representation (i.e., a fragment of first-order logic without negation) offers sufficient descriptive power for our task, rather than frameworks designed to deal with the richness of natural language, such as File Change Semantics (Heim, 1982) or Discourse Representation Theory (Kamp, 1984), which are equivalent to first-order logic.

Existing representation schemes used for semantic web applications (based on RDF) are not entirely suitable for our purpose either. The format used in the FOAF (Friend of a Friend) ontology is suitable for describing social relations between persons (Ding et al., 2005), but would need considerable extensions to describe entities on tombstones such as biblical references, symbolism and events. In event ontologies, such as the Simple Event Model (Van Hage et al., 2011), events take centre stage, and even though there are two events that play key roles in tombstone descriptions (birth and death), it would be cumbersome to describe these ‘default’ events in the format required in these frameworks. Nonetheless, our proposed representation for these two core events found on tombstone descriptions allows for a straightforward mapping to standard event ontologies. Likewise, the Schema.org initiative (Guha et al., 2016) lacks coverage of the specific attributes of tombstones and is therefore not directly suitable for our needs. However, it is consistent with the framework that we propose and future work could provide an extension of Schema.org to include tombstones. Existing formal ontologies defined for applications in the digital humanities, such as CIDOC-CRM (Biagetti, 2016), come very close to our goals but are not tailored to tombstone inscriptions specifically and are therefore not suitable for modelling all of the details required for a good characterisation of the information found on tombstones.

Following this line of reasoning, and in accordance with recent developments in computational linguistics, our choice for representing tombstone interpretations is the PENMAN notation, a representation originally introduced for natural language generation systems (Kasper, 1989; Bateman et al., 1989; Bateman & Paris, 1989; Bateman, 1990), also known under the names of Penman Interface Notation or SPL (Sentence Plan Language). This representation format is flexible (it can be visualised in various ways), is relatively easy to understand by human annotators (Banarescu et al., 2013), is supported by various software packages for evaluation purposes (Cai & Knight, 2013), and several semantic parsers have been developed for it (see Section 5). An example of this kind of representation is shown in Fig. 3.

SPL describes entities and relations between them. Every entity is assigned a unique identifier and needs to be declared as an instance of a concept taken from a certain ontology. This is done using the slash notation. For instance, in Fig. 3, p2 is an entity of the concept `male.n.02`, coded as `(p2 / male.n.02 ...)`, where the dots are used as a placeholder to add more information about p2. Here, p2 is an arbitrarily chosen identifier that has not been used before (any identifier could have been chosen that had not been used before in the same meaning representation). For clarity, however, we use certain notational conventions in the annotations, and identifiers are coded as a single lower-case letter followed by an integer (e.g., p1, p2, r1, r2, s1, s2, and so on). Constants are written in double quotes. Examples



```
(t00196 / tombstone.n.01
  :sym (s1 / palm.n.03)
  :ent (p1 / female.n.02
    :rol (r1 / wife.n.01
      :tgt p2)
    :rol (r2 / mother.n.01)
    :nam "TJITSKE DE VRIES"
    :dob "1884-02-29"
    :dod "1933-05-01"
    :rol (r3 / spouse.n.01
      :tgt p2))
  :ent (p2 / male.n.02
    :nam "JAN HUISMAN"))
```

Fig. 3 Tombstone (left) with annotated meaning representation in SPL (right)

of constants in Fig. 3 are the dates and the proper names of people mentioned on the stone. The meaning representations can also be visualised as a directed graph, as shown in Fig. 4.

Moreover, the information can be simply stored as a collection of database triples, making it an ideal format to link with other knowledge resources or relational databases on the semantic web. Figure 5 illustrates this idea, as well as some SPARQL (Arenas et al., 2018) example queries that could be used to search for tombstone images with specific attributes.

The SPL formalism gives us a general representation format, corresponding to a fragment of first-order logic (see Appendix B). What SPL does not specify is the

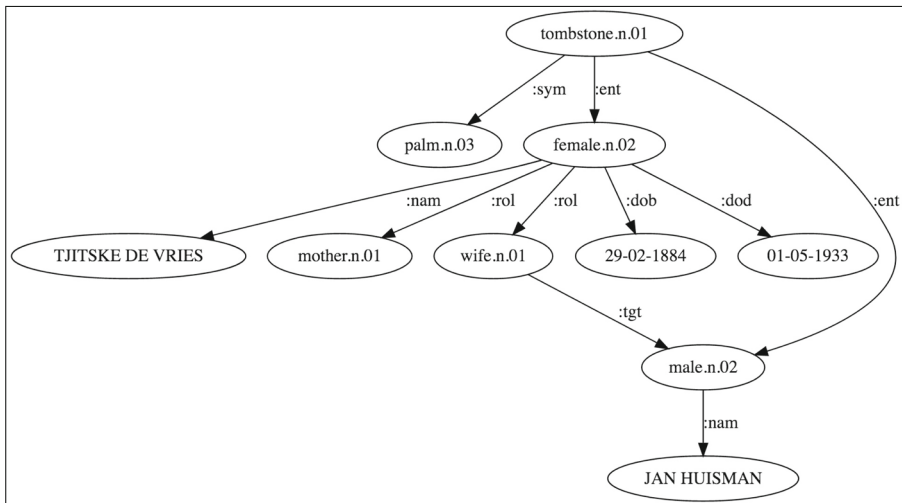


Fig. 4 Graph visualisation of the tombstone information of Fig. 3


```

t00196 :isa tombstone.n.01          SELECT ?t
t00196 :sym s1                      WHERE { ?t :isa tombstone.n.01 .
s1      :isa palm.n.03              ?t :ent ?p .
t00196 :ent p1                      ?p :isa female.n.01 }
p1      :isa female.n.02
p1      :rol r1                      SELECT ?t ?d
r1      :isa wife.n.01              WHERE { ?t :isa tombstone.n.01
r1      :tgt p2                      ?t :ent ?p .
p1      :rol r2                      ?p :dod ?d .
r2      :isa mother.n.01            (FILTER ?d < "1900-01-01") }
p1      :nam "TJITSKE DE VRIES"
p1      :dob "1884-02-29"           SELECT ?s
p1      :dod "1933-05-01"           WHERE { ?t :isa tombstone.n.01 .
p1      :rol r3                      ?t :sym ?s .
r3      :isa spouse.n.01            ?t :ent ?p .
r3      :tgt p2                      ?p :rol ?r .
t00196 :ent p2                      ?r :isa widow.n.01 }
p2      :isa male.n.02
p2      :nam "JAN HUISMAN"

```

Fig. 5 Representation of the tombstone information of Fig. 3 as RDF-like triples (left) and three example SPARQL queries (right): the first selects tombstone inscriptions with female people, the second selects tombstones with people who passed away before the 20th century, and the third returns symbols on stones with widows mentioned on them

vocabulary of terms: the concepts and relations that are used for a specific domain or task. For instance, in the Abstract Meaning Representation project (Banarescu et al., 2013), SPL is used in combination with relations adopted from PropBank (Palmer et al., 2005) and concepts taken from OntoNotes (Hovy et al., 2006). In our work, we populate our vocabulary with concepts defined in WordNet (Section 3.2). WordNet (Fellbaum, 1998) is a lexical resource used extensively in computational linguistics and information retrieval, and is converted to standard ontologies (Gangemi et al., 2003) and RDF (Van Assem et al., 2006). In addition, we specify a novel set of relations tailored to the task of tombstone inscription interpretation in order to produce compact meaning representations (Section 3.3).

3.2 Encoding concepts: WordNet

There are two kinds of entities in SPL: constants and variables. Constants (literals) represent concrete values and are either numbers or strings. In our variant of SPL, we enclose strings in double quotes and transform them to upper case for normalisation purposes. Numbers are never enclosed in quotes, unless they are used as names. If an entity is represented by a variable, it needs to be linked to a concept. In our variant of SPL, we adopt WordNet synonym sets (synsets) as concepts (Fellbaum, 1998). A synset is a set of lemmata that, when used in an appropriate context, have the same or similar meaning. Ambiguous words are therefore binned in two or more different synsets.

A synset can be referred to by its identifier or by one of its members and a sense number. For notational convenience, we choose the latter option and use the lemma – part-of-speech – sense format familiar in natural language processing. For instance, `oak.n.01` is the concept for “the hard durable wood of any oak, used especially for furniture and flooring”, whereas `oak.n.02` is the concept for “a deciduous tree

of the genus *Quercus*; has acorns and lobed leaves”. Here, *n* stands for the part-of-speech noun, and the numbers refer to the first and second sense of the noun “oak”. The other parts of speech are *v* (verb), *a* (adjective) and *r* (adverb).

WordNet is organised by relations between synsets. The hypernym relation is especially useful for implementing a semantic search, because it gives us the means to assign specific as well as general concepts to objects. Some examples are given in Appendix C of how to exploit WordNet relations for the purpose of producing background knowledge.

English WordNet, a lexical database incorporating an ontology (Fellbaum, 1998), was chosen for its large coverage. It turns out that we can use the same base of concepts for describing persons, locations, jobs, family roles, symbols, religious references, and more. Even though we make use of the English WordNet, extension to multilingual WordNets (Bond & Foster, 2013) is certainly an idea worth considering. It is, however, important to realise that the meaning representations show concepts, not English words, and although these concepts are triggered by words, we assume some language-independence here. It could be the case that extensions are required in the future, as some non-English words could be untranslatable into English directly. Nonetheless, the original WordNet causes no substantial limitations for our current undertaking.

3.3 Encoding relations

The WordNet concepts provide an ontology of entities. However, we also need a vocabulary of relations to connect these entities. It is possible to re-use existing general purpose ontologies and adopt the event-based thematic role from PropBank (Palmer et al., 2005), VerbNet (Kipper et al., 2008) or FrameNet (Baker et al., 1998). Here, instead, we design our own set of relations, tailored to the task of describing tombstone texts, to keep the resulting representations compact.

A relation is signalled by a colon in SPL. In the example in Fig. 3, we have several relations, such as `:sym`, `:ent`, `:nam`, `:rol` and `:dob`. An entity can be linked to one or more entities. For instance, in Fig. 3, the entity `p1` is related to entity `r1` by the `:rol` relation, and related to the constant `"01-05-1933"` by the `:dob` relation. This gives us domain-specific relations such as `:dob` in Fig. 3, rather than the more cumbersome “X died in event E, and E took place on date D”. A link to general purpose relations can still be provided by postulating a set of axioms (see Appendix C).

The general idea of a meaning representation is that it not only shows the entities involved but also how they relate to each other. In our domain of gravestone inscriptions, we ensure that the resulting meaning representation satisfies the property of connectivity with the help of the `:ent` and `:sym` relations that relate the tombstone to the entities that it displays. As a tombstone could comprise information on more than one deceased person without supplying information on how these people relate to each other, the root of the meaning representation is always the tombstone itself, as the example in Fig. 4 illustrates.

SPL has a feature that makes it possible to inverse relations (Bateman, 1990). An inverse relation is indicated with the `-of` suffix. Take, for instance, the meaning

representation (`b1 / book.n.01 :prt (c1 / chapter.n.01)`), which says that a book `b1` (the root) has a chapter `c1`. We can bring the chapter to the foreground by inverting the relation, arriving at (`c1 / chapter.n.01 :prt-of (b1 / book.n.01)`), where now the root is taken by `c1`, a chapter of book `b1`. It is important to note that these two meaning representations are equivalent, but are just packaged differently. This ability is sometimes useful for annotation purposes. We use inverse relations for biblical references, geographical relations and comparative constructions.

The complete collection of relations (and their definitions) is presented in Table 1, covering 32 different relation types. The usages of these relations and how they relate to concepts are explained in Section 4, where we discuss the phenomenon found in the inscriptions.

4 A formal description of tombstone phenomena

The number of types of entities found in the tomb inscriptions of our collection is, although limited, perhaps more than would be expected, and comprises people, dates, locations, occupations, organisations, family roles, measures, quotes (epitaphs) and references. Besides textual entities, objects are also displayed with the use of graphical images.

In this section, we describe the challenges to represent these entities in a formal way, by summarising the variety in which they appear on tombstones, how they can be linked to an ontology and, for computational purposes, how they can be represented in a normalised way. We cover nearly all of the types of information found on tombstones in our dataset, with the exception of epitaphs and cause of death. Epitaphs pose challenges because of the variety that they introduce and are perhaps less suitable to be included in a meaning representation. Cause of death is rarely mentioned on the tombstones in our dataset (less than 0.1%) and therefore represents a sample that is too small on which to design and test a proper meaning representation.

4.1 People

The names of the deceased or the commemorator (sometimes both) may appear on tombstones. On rare occasions (less than 1% in our dataset), the designer of the stone is also mentioned, although usually in a less prevalent position. Full proper names are usually used for the deceased, although sometimes only first names appear (especially for children) or surnames with initials or abbreviations of given names, and occasionally titles or patronymic suffixes are displayed as well. For married women, maiden names are regularly provided, and alternative names or nicknames can sometimes be found as well. Only on very rare occasions are people left unnamed (in 24 cases). Furthermore, people can also be implicitly introduced by the use of possessive pronouns (“my husband”, “our aunt”).

Named people are linked to a stone with the `:ent` relation. Usually (in 95% of the cases), the perceived gender of the person can be determined on the basis of the

Table 1 Relations used in the tomb meaning representations

Relation	Meaning
:ent	X (gravestone) has entity Y (person)
:sym	X (gravestone) has symbol Y (symbol)
:ref	X (gravestone) has (biblical) reference Y (text)
:lan	X (literal) is written in language Y (ISO 639-2 language code)
:nam	X (person/location) has name Y (string)
:pfx	X (person/location) has prefix Y (string)
:sfx	X (person/location) has suffix Y (string)
:dod	X (person) has date of death Y (date)
:dob	X (person) has date of birth Y (date)
:pod	X (person) has place of death Y (location)
:pob	X (person) has place of birth Y (location)
:occ	X (person) has occupation Y (occupation)
:rol	X (person) has family role Y (role)
:age	X (person) has age Y (measure)
:loc	X (occupation) has location Y (location)
:org	X (occupation) has organisation Y (string)
:hco	X (occupation) has hisco code (code)
:mem	X (group) has member Y (entity)
:qua	X (group) has quantity Y (number)
:sub	X (group) has subgroup Y (group)
:prt	X (composite) has part Y (entity)
:op1	X (comparison) has first operand Y (measure)
:op2	X (comparison) has second operand Y (measure)
:ord	X (entity) has ordinality (number)
:att	X (entity) has attribute Y (attribute)
:tgt	X (role) has owner/possessor Y (entity)
:beg	X (role) has beginning Y (date)
:end	X (role) has ending Y (date)
:dur	X (role) has duration Y (measure)
:aft	X (role) played role after Y did (role)
:bef	X (role) played role before Y did (role)
:geo	X (location) has GeoNames identifier Y (literal)

first name, maiden name, title or family relations provided on the stone. The WordNet concept *female.n.02* is used for female persons and *male.n.02* for male persons, without suggesting that these are the only options for representing gender. In cases where the gender identity of the person is non-binary or other-gendered, or in case the gender of the person cannot be recovered on the basis of the information provided by the inscription, the gender-neutral *person.n.01* is chosen. For

names on family graves, the concept `family.n.04` is used. Groups of persons are coded with the WordNet concept `people.n.01`, meaning “a group of human beings”.

Names are represented as constants; that is, in upper-cased strings in double quotes. Abbreviations occur frequently in names, so the annotation guidelines should be clear on when to expand or not. Initials are not expanded because the ability to do this depends heavily on external resources (e.g., online family trees); multiple initial letters are encoded without spaces between punctuated letters. Abbreviated prepositions in the family name (common in Dutch surnames) are expanded, because these are mostly unambiguous and add another level of normalisation in the meaning representation.

If more than one name is used for the same person, and they are different (for instance in the case of nicknames), an entity will be associated with more than one name via multiple use of the `:nam` relation. However, if there are several names on a stone for the same person, only the most specific name is coded in the meaning representation (because the less specific name is subsumed by the more general one, following basic principles of semantics). Pronouns are sometimes used in tomb inscriptions to establish coreference between named entities. These are not only possessive pronouns, but also the pronoun “both” (Dutch *beiden*).

Patronymic suffixes, honorifics and titles are not considered to be part of the proper name itself and introduce their own relations, either via `:pfx` (prefix) or `:sfx` (suffix). Examples of name prefixes are honorifics or titles (paraphrased in English): IR (engineer), JNKV (lady), DS (minister), MEVROUW (Mrs), HEER (Mr), DR (doctor). Examples of name suffixes are AZ (Arent’s son), B.ZOON (Berend’s son), EZ (Egbert’s son or Evert’s son), GZN (Geert’s son), JZ or JZN (Jan’s son), HZ or HZN (Hendrik’s son), HARMZN (Harm’s son), PZ (Pieter’s son), KZ (Koop’s son), RZ or RZN (Roelof’s son), WZN (Willem’s son). These examples show that, even though abbreviated prefixes are generally resolvable, patronymic suffixes require careful study of genealogical databases. We therefore choose not to expand suffixes, and for consistency we do not expand prefixes either.

The representation of names that we propose is incredibly simple. It is possible to analyse names further by breaking them down into given names, family names, maiden names, nicknames, and perhaps more. This is not simple or straightforward in some cases without the addition of external knowledge resources, and the solution we adopt has the advantage of being applicable to many cultures and traditions of naming people. Future work could address a more detailed semantic analysis of personal names.

4.2 Date expressions

An essential piece of information on a gravestone is the date of death (although family graves commonly do not contain dates). In most cases, both the date of birth and date of death are shown (in that order), often with a textual (Fig. 3) or symbolic (Fig. 6) sign that indicates birth or death. The three components of a date (the day, the month and the year) are usually all present and, when they are, they come in this order in our dataset. In some cases (1.3%), only the year is provided.



```
(t1 / tombstone.n.01
: sym (s1 / branch.n.01)
: ent (p1 / male.n.02
: nam "HENDRIK VAN DAM"
: dob "1873-01-01"
: pob (c1 / town.n.01
: nam "HAREN"
: geo "2754841")
: dod "1947-10-23"
: pod c1
: ref (v1 / verse.n.01
: ord 1
: att (a1 / rhymed.a.01)
: prt-of (c2 / chapter.n.01
: ord 119
: prt-of (b1 / Psalm.n.01))))))
```

Fig. 6 Tombstone with biblical reference (left) and corresponding meaning representation (right)

The day of the month and the year are nearly always written in numbers, although the century may sometimes be omitted (e.g., '42 for 1942). The expressions denoting months show a large variety, ranging from simple numbers to Roman numerals, full names (sometimes with different spellings) in lower-case or upper-case characters and abbreviations in various forms. There is also a noticeable heterogeneity in the use of punctuation in date expressions, with short, long and lowered hyphens, colons, periods and diverse spacing. An impression of this variety is given in Table 2.

Table 2 Different formats of date expressions found on tombstones in Groningen with their normalisation

Original Expression	Normalised Expression
'44	1944-XX-XX
1932	1932-XX-XX
25 09 1925	1925-09-25
12-10-33	1933-10-12
5-2-1910	1910-02-05
7.12.'70	1870-12-07
24 _6 _ 1873	1873-06-24
21-3-1887	1887-03-21
14 — 2 — 1916	1916-02-14
4.6.1891	1891-06-04
7 . 9 . 1900	1900-09-07
19 XII 1924	1924-12-19
27 JAN:1866	1866-01-27
16 JANR.1849	1849-01-16
20 Feb. 1767	1767-02-20
24 FEBR. 1899	1899-02-24
28 Febr.1896	1896-02-28
8 Maart 1955	1955-03-08
1 MAART 1983	1983-03-01
31 MRT. 1931	1931-03-31
8 Junij 1871	1871-06-08
8 AUG ^S 1935	1935-08-08

All date expressions are normalised using the format "YYYY-MM-DD", where DD are digits for the day of month (01–31), MM digits for the month (01–12) and YYYY digits for the year. In case the days and months are not provided, "XX" is used to represent this information. In partial dates where the century is not provided, contextual information (the state of the tombstone, general knowledge of life expectancy) or online genealogy databases are used to complete the missing information.

We adopt this compact notation of date expressions for practical reasons, as it will ease the process of annotation. However, for the purpose of temporal inference (see Appendix C) or fine-grained evaluation, we may want to deconstruct this notation into its atomic parts. For instance, the 1st of February 2003 is represented in our default scheme by the constant "2003-02-01". It is straightforward to break this down automatically into the structured, but equivalent (`d1 / date.n.05 :dom "01" :moy "02" :yoc "2003"`), where the relations `:dom`, `:moy`, `:yoc` designate the day of month, month of year, and year of century, respectively.

4.3 Toponyms

When toponyms (names of villages, towns or cities) are used in our dataset, they commonly denote the birthplace or the place of death. Location names are occasionally (4%) used in connection with the deceased's line of work. Note that, on tombstones of different cultures, toponyms can also be used to indicate the origin of settlers or social identity, as displayed on Taiwanese tombstones (Streiter et al., 2010b; Streiter et al., 2008). Besides the name of a village or town, further information is sometimes given for non-local places or ambiguous place names, such as the region, province or country. Many tombstone descriptions contain location names. In our dataset, we find an average of one mentioned location per tombstone, but if a location name is found on a stone, it is usually more than one (in 70% of the cases).

Interpreting toponyms raises the challenge of grounding; in other words, of mapping to a location in the real world (Leidner, 2008; Kew et al., 2019). Names are sometimes abbreviated, or are written following a non-standard or old-fashioned spelling, and this can make them hard to interpret (Table 3). An additional problem is that many names for locations are ambiguous with respect to their reference in the real world. Villages that go by the same name are widespread in the Netherlands (some examples from our dataset are Den Oever, Erica, Hoorn, Langelo, Nes, Niek-erk, Noordwolde, Oostwold, Spijk, Siegerswoude, Westerlee, Winsum, Zevenhuizen and Zuidwolde) or can be used in other countries (Kampen, Oostburg, Roden, Zutphen). Another type of referential ambiguity is caused by cities, municipalities or provinces that share the same name (Groningen, Grootegast, Vries).

Our approach for toponym interpretation is similar to that of Leidner (2008), who grounds toponyms with an extensional semantics in the form of a pair of latitude and longitude of the centroid of the location. However, instead of using geographical coordinates, we consult the geographical database GeoNames (Maltese & Farazi, 2013) for a unique identifier for a given place name. GeoNames is a widely-used multilingual gazetteer in the semantic web (Ahlers, 2013) that makes it possible to link places to geographical coordinates and further topographical data. It also contains historical names, which is convenient as some villages found on old tombstones no

Table 3 Examples of normalisation of toponyms with their GeoNames identifiers

Original Expression	Normalised Expression	GeoNames	Coordinates
ZUTFEN	ZUTPHEN	2743608	52 8 0 N 6 12 0 E
Tinaarlo	TYNAARLO	2746038	53 4 36 N 6 37 3 E
GRON.	GRONINGEN	2755251	53 13 0 N 6 34 0 E
's-Gravenhage	DEN HAAG	2747373	52 5 44 N 4 18 59 E
Den Bosch	'S-HERTOGENBOSCH	2747351	51 42 36 N 5 18 41 E
H-Opeinde	HARKEMA-OPEINDE	2754821	53 11 8 N 6 8 12 E
Neêrlands Indië	NEDERLANDS-INDIË	1643084	2 0 0 S 118 0 0 E
Bovenknijpe	DE KNIPE	2758492	52 58 7 N 5 58 16 E
BOXUM	BOKSUM	2758458	53 10 38 N 5 43 47 E
LUTJEPOST	LUTKEPOST	2751365	53 14 28 N 6 9 32 E
De Leek	LEEK	2751834	53 9 42 N 6 23 5 E
UITH.MEEDEN	UITHUIZERMEEDEN	2745968	53 13 0 N 6 34 0 E
ANNERV.-KAN.	ANNERVEENSCHekANAAL	2759724	53 4 40 N 6 48 5 E

longer exist or have been incorporated into larger urban districts.² Example mappings from toponyms to GeoNames identifiers are shown in Table 3.

As with names of persons, location names are orthographically represented by upper-cased strings. Names of locations may differ across languages, and it would be counter-intuitive to translate names in a particular language for the sole purpose of normalisation. Instead, we preserve the original name and annotate the language using the three-letter codes proposed in the ISO 639-2 standard. A full example in SPL for a toponym (the Dutch-Frisian city of Leeuwarden) is the following:

```
(c1 / city.n.01
  :nam (n1 / name.n.01
    :lit "LJOUWERT"
    :lan "FRY")
  :geo "2751792")
```

As this example shows, we also associate every toponym with a WordNet concept that uniquely identifies its type. Table 4 shows concepts that we use in our domain, together with their definitions adapted from WordNet.

A complete SPL representation with toponyms is shown in Fig. 6. This figure also demonstrates the shorthand notation that we use in SPL for Dutch place names (the majority), because we only encode the language if it is not Dutch, to make

²In our dataset with 1,100 tombstones, we found seven location names that were not listed in GeoNames: “Bargermeer” (a hamlet), “Boxum” (but its synonym “Boksum” was present), “Holslootbrug” (a bridge), “Leeuwenburg” (a farmhouse), “Lutke Saaxum” (a hamlet), “Menneweer” (another hamlet) and “Nijhuzum”, Frisian for “Nijhuizum”).

Table 4 Types of toponyms, defined by WordNet concepts

WordNet concept	Definition
hamlet.n.03	a settlement smaller than a village
village.n.02	a settlement smaller than a town
town.n.01	an urban area smaller than a city
city.n.01	a large and densely populated urban area
municipality.n.01	an urban district
province.n.01	the territory occupied by an administrative district
state.n.01	the territory occupied by an administrative district
country.n.02	the territory occupied by a nation
island.n.01	a land mass surrounded by water
ocean.n.01	a large body of water
colony.n.05	a geographical area politically controlled by a distant country

the annotation task lighter for the human annotator. Hence, the shorthand expression :nam "LEEWARDEN" is interpreted as :nam (n1 / name.n.01 :lit "LEEWARDEN" :lan "NLD").

Finally, we found two types of locative demonstrative pronouns in our dataset that are used to refer to locations. The Dutch word *aldaar* ('the earlier mentioned place') is regularly used for the place of death in case it is the same as the birthplace of the deceased (in our dataset, it is never utilised to refer to a location mentioned in connection with another deceased person). Because it refers back to an already mentioned place name, an earlier introduced referent in the SPL representation is reused, as Fig. 6 shows. Sometimes, the Dutch word *alhier* ('here, at this place') is used to refer to the village, town or city in which the cemetery is situated. This information is impossible to retrieve from the text on the stone, but the metadata of the stone's image contains the GPS coordinates of the location, which is then used to determine the place name.

4.4 Family relationships

Family relationships (such as wife, husband, mother, father, grandmother, grandfather, great-grandmother, great-grandfather, widow, widower, parent, child, grandparent, grandchild, daughter, son, spouse, aunt, uncle, sister, brother, and rarer ones like grand-aunt, grand-uncle, daughter-in-law, brother-in-law, sister-in-law, foster-parent and housemate) are often displayed on tombstones. We propose to represent them as WordNet concepts. However, as they express a relationship between two persons (or groups of persons), we also need to have the means to signify the person that carries out the role of the relation, and the person that is the target. These links are made via the :rol and :tgt relations.

Several instances of family roles are shown on the tombstone introduced earlier in Fig. 3. Three roles are supplied for person p1: she plays the role of wife.n.01

(“LIEVE VROUW”), `mother.n.01` (“MOEDER”) and `spouse.n.01` (“ECHT-GENOOTE”). As the meaning representation for this stone shows, roles introduce their own entity variable; in this case, we have `r1`, `r2` and `r3`. The first two roles do not explicitly express their target on the inscription. However, the target of the first role, that of wife, can be derived from further information on the stone, and is denoted by `p2`. We therefore end up with `:rol (r1 / wife.n.01 :tgt p2)`. The target of the mother role is left unspecified – the stone provides no information. The target of the third role of spouse is explicit, it is again `p2`. Note that, although the meaning representation contains redundant information, as `wife.n.01` subsumes `spouse.n.01`, we opt to represent both family relationships for reasons of consistency (see Appendix C for notes on inferences).

The motivation to have explicit entities for roles is that a person can play several roles – even the same type of role with different targets – and that roles can be anchored in time. For example, a person can first be a widower of one person, and later a widower of another person. This example also brings up an interesting characteristic of roles, which is the time span that they implicitly denote. This is sometimes made explicit, and to connect a role to a measurement phrase we use the `:dur` relation conveying duration. If several roles are mentioned, they can also be compared to each other, as in “first widow of X, later of Y”. To express these temporal relations, we use `:bef` (before) and `:aft` (after) as relations between roles.

4.5 Phrases of measurement

Tombstones sometimes provide information about the age of the deceased or about the length of a certain state of relationship. This phenomenon is relatively rare in our dataset: nearly 4% of tombstones have descriptions of measurements. We use the WordNet concept `measure.n.02` (“how much there is or how many there are of something that you can quantify”) as a group with members that denote a certain unit of measurement (see Appendix D for further information on the semantic interpretation of groups) and quantity that specifies the number of units that have been measured.

Measurement phrases on stones denote periods and are measured in years, and sometimes (in 10% of the cases) in finer units of time (months or days). In our dataset, we capture these with the following WordNet concepts for units of measurement: `year.n.01`, `month.n.01` and `day.n.01`. Quantities are described by numbers and linked to the measurement with the `:qua` relation. Numbers, as usual in formal systems, are constants, but for notational simplicity we do not enclose them in double quotes. For instance, the expression “75 years” is represented as a measure comprising 75 members of the unit of measurement `year`:

```
(m1 / measure.n.02
  :qua 75
  :mem (u1 / year.n.01))
```

We introduce two relations to link measurements to other entities: `:age` to specify the age of people and `:dur` to specify the duration of relationships (see Table 1).

Moreover, measurement phrases can be combined with vague modifiers, and global units are sometimes joined with finer units of measurement to add precision to the phrase.

First consider the latter. This concerns measurement phrases that combine different units of measurement, as in the phrase “57 years and four months”. The problem here is that there is some summation taking place of the two measures expressed, rather than viewing two independent phrases separately. One solution would be to normalise the two different units to a common unit and sum the quantities. In the example above, this would yield a measure of units of $57 \cdot 12 + 4 = 688$ months. However, the meaning representation would deviate considerably from the surface string. An alternative, and the solution that we propose, is to represent complex measures in a recursive way, using sub-groups:

```
(m1 / measure.n.02
  :sub (m2 / measure.n.02
        :qua 57
        :mem (u1 / year.n.01)))
  :sub (m3 / measure.n.02
        :qua 4
        :mem (u2 / month.n.01)))
```

Measurements can also be modified with expressions that introduce some kind of vagueness. Two different types of modification for measurements are encountered in our dataset: the Dutch adverbs *ruim* (over) and *bijna* (nearly). The meaning of *ruim*, as in “ruim 40 jaar”, indicates an amount more than the concrete value specified by the phrase that it modifies, but its precise meaning is hard to pinpoint. It seems to carry a conventional implicature (Grice (1975)) too – in our example it conveys that the age is less than, but close to, 41 years. Its antonym, *bijna*, as in “bijna 54 jaar”, denotes an amount *less* than that expressed by the phrase it modifies. Again, it triggers a conventional implicature (namely that the amount that is lacking to fulfil the specified amount is very small and less than one unit of it).

A simple, practical solution is to ignore this type of vague modification. As *ruim* $N U$ (where N is a number and U a unit of measurement) entails $N U$, we could just ignore the contribution of the modifier in our semantic analysis. Likewise, for *bijna* $N U$, we could represent its meaning as $N-1 U$. However, this strategy could have undesired consequences (for instance, consider the case where $N = 1$) and places little faith in our journey of formalisation. The direction that we take is to make these modifiers explicit in the meaning representations, by introducing an additional measure entity and relating it to the entity that it modifies. We then make a comparison between these two measures, and specify whether the first is more or less than the second. For instance, “more than 40” is mapped to the following representation:

```
(c1 / more.r.01
  :op1 (m1 / measure.n.02)
  :op2 (m2 / measure.n.02
        :qua 40
        :mem (u1 / year.n.01)))
```

Here we have the WordNet concept `more.r.01` (and likewise, we adopt the complementary `less.r.01`), which is related to the comparator via `:op1` and the comparand via `:op2`. We apply relation inversion (see Section 3.3) to make the measure specified by the comparator the root of the meaning representation:

```
(m1 / measure.n.02
  :op1-of (c1 / more.r.01
    :op2 (m2 / measure.n.02
      :qua 40
      :mem (u1 / year.n.01))))
```

This representational treatment of vague modifiers gives us a first handle on its semantic impact, but future work, including empirical research (i.e., looking at more data), is required to fully pinpoint the meaning of this interesting linguistic phenomenon.

4.6 Occupations and organisations

Occasionally, information about the line of work in the life of the deceased is provided on a tombstone (in about 8% of the stone inscriptions in our dataset). We integrate this information into our meaning representation for stone inscriptions by combining an existing classification of job descriptions (HISCO) with WordNet concepts (Table 5). HISCO, short for Historical International Standard Classification of Occupations, is a multilingual resource for classifying occupations (Van Leeuwen et al., 2002; Tobin et al., 2019), and it provides a hierarchical way of organising job titles. This is accomplished with a single code comprising five digits, where the first digit indicates a general class, and the digits following make the occupation more specific (e.g., “0” represents generic technical workers, “04” officers of ships and aircraft, “041” aircraft pilots, and “04120” air transport pilots). HISCO contains examples of occupations in several languages, including Dutch, and is therefore especially suitable for our domain of interest.

Occupations are linked to the person with the `:occ` (has occupation) relation. The occupation itself is represented by a WordNet concept and is connected to a HISCO code with the relation `:hco`. For instance, the job of *goudsmid* is represented as `:occ (o1 / goldsmith.n.01 :hco "88050")`.

In some cases, further information about the job is given on stone inscriptions: the period or duration of the occupation, the organisation that the person worked for, or the location of the workplace. A period is indicated with the `:beg` and `:end` relations. Locations (toponyms) are linked to the occupation using the `:loc` relation. Organisations are represented in a similar way as people: as named entities combining an appropriate WordNet concept for the type of organisation with its name. Organisations comprise, for instance, companies (`company.n.01`), schools (`school.n.01`), unions (`union.n.07`), religious communities (`community.n.02`) and universities (`university.n.02`). They are linked to the occupation using the `:org` relation.

Table 5 Occupations found on Dutch tombstones, and corresponding WordNet concepts and HISCO code

Dutch original	WordNet concept	HISCO
scheepskapitein	skipper.n.02	04217
zeeloods	pilot.n.02	04240
bioloog	biologist.n.01	05110
arts, geneesheer	doctor.n.01	06105
tandarts	dentist.n.01	06310
apotheker	pharmacist.n.01	06710
wijkverpleegster	visiting_nurse.n.01	07130
kantonrechter	judge.n.01	12210
notaris	notary.n.01	12310
onderwijzer, onderwijzeres, leraar, hoofdonderwijzer	teacher.n.01	13320
hoofd der school	principal.n.02	13940
kerkvoogd, ouderling	elder.n.03	14000
predikant	pastor.n.01	14120
dominee	preacher.n.01	14120
auteur	author.n.01	15120
beeldhouwer	sculptor.n.01	16120
kunstschilder, zeeschilder	painter.n.01	16130
socioloog	sociologist.n.01	19220
burgemeester	mayor.n.01	20110
wethouder	alderman.n.01	20110
ontvanger, gemeenteontvanger	taxman.n.01	31020
assistent posterijen	postal_clerk.n.01	37020
rijksveldwachter	constable.n.03	58220
wachtmeester	sergeant.n.02	58220
soldaat, militair	soldier.n.01	58340
bewaarder	keeper.n.02	58940
landbouwer	farmer.n.01	61220
goudsmid	goldsmith.n.01	88050
steenfabrikant	manufacturer.n.02	89242
vioolbouwer	luthier.n.01	94120
machinist	engineer.n.02	98320

4.7 Biblical references and quotes

Quotes from the Bible, or short references to these texts, appear on around 15% of the tombstones in our dataset. They are meant as a source of comfort and relief for the commemorators, and their use is typical on Protestant gravestones (Krommenhoek, 2016). The references are highly structured, referring to one or more verses of a book in the Old or New Testament. For instance, “Ps.119:1 BER” refers to the rhymed version of the first verse of the 119th book of Psalms. There are three ways in which

references appear on stones: a reference without the text itself; the text without its reference; or a combination of reference and text. In the first case, we capture just the reference in the meaning representation (because various versions of the text exist, so it is impossible to tell which one was meant). In the second and third case, we represent both the reference and the quote.

Our annotation makes the structure of biblical references explicit by distinguishing between verses (*verse.n.01*), chapters (*chapter.n.01*) and books. Sometimes, just a part of a verse is referred to, in which case we use the concept *part.n.09*. All of the books of the Bible are listed as instances of the Old or New Testament in WordNet, so there is no need to include another ontology to uniquely identify biblical references. An example of the resulting representation is shown in Fig. 6.

Psalms have been adapted to rhymed versions to make them more suitable for congregational singing (Slenk, 1969), and whether or not this process of conversion is applied is sometimes explicitly indicated on tombstones. This is accomplished by a variety of possible expressions, including “BERIJMD”, “BER.”, “BER” or “berijmdt” (rhymed) and “ONB.” or “O.B.” (unrhymed). We represent these attributes with the aid of the *:att* relation and the WordNet concepts *rhymed.a.01* and *unrhymed.a.01*. If more than one verse is cited, as in “PSALM 118:17-20”, we use the group construction (see Appendix D).

We adopt the relation *:rel* to link the biblical reference to other entities of the tombstone inscription. However, it is not always clear to which entity this link should be made. On stones with more than one deceased, biblical references are sometimes clearly positioned to apply to a particular person. In such cases, it is obvious that the intention of the commemorators was to connect specific references to each of the deceased. If there is more than one deceased person but only a single biblical reference (placed at the bottom of a stone), the connection is less clear. In this case, the reference is linked to the stone itself.

4.8 Symbolism

Tombstones are often (in more than half of the cases in our dataset) decorated with interesting symbols – some of them modern, others older. Our primary aim is not to understand the symbolism itself by deciphering its deeper meaning, even though this can give information about the beliefs and social status of the deceased (Krommenhoek, 2016): the Christian cross is a symbol of Catholicism, the Star of David of Judaism, the Square and Compasses of Freemasonry, and a crest of the Salvation Army.

Rather, we take a descriptive approach, capturing the symbols that appear on stones with concepts, and leaving further interpretation for other branches of research. An existing classification system designed for art is ICONCLASS (Coupric, 1983). This is used by researchers (mostly art historians) to annotate an image with the concepts it visualises. Its definitions are hierarchically ordered, in a way similar to the HISCO scheme for occupations (Section 4.6). The inclusion of ICONCLASS in our meaning representation could be realised with a relation similar to *:hco*, but it requires expert knowledge from an art historian and is therefore the subject of future

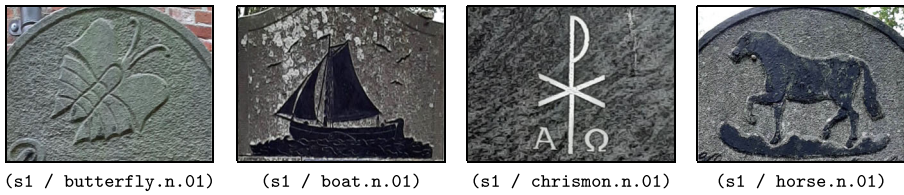


Fig. 7 Examples of symbols found in our datasets with their corresponding WordNet concepts

interdisciplinary work. Instead, we use WordNet to describe the symbols found on the tombstones in our dataset (Fig. 7).³

The semantic relation that we use to indicate that a stone is decorated with symbols is `:sym` (see Table 1). We distinguish two main types of symbol representations: single symbols and composite symbols. Symbols that appear alone are simply represented with an appropriate WordNet concept that is as specific as possible, examples of which are given in Fig. 7. Some symbols are displayed with prominent attributes: a broken flower, a winged sandglass, an inverted torch or clasped hands. These are represented with the help of the `:att` relation, as Fig. 8 demonstrates, with the corresponding WordNet concept for adjectives to capture the type of modification.

Composite symbols are two or more symbols that are visualised next to each other or touching each other. Examples are two fronds that are tied together with a ribbon, a weeping willow decorated with a scythe and an inverted torch, a dove with a branch in its beak, a horse standing next to a gate, or a hand holding a torch. We represent these compound symbols with the WordNet concept `composite.n.01` and the relation `:prt`. It goes beyond the scope of this article to describe composite symbols with spatial relations and the position of the symbols on the gravestone. However, the meaning representation that we develop here could accommodate such an extension.

5 Automatic reading of tombstones

To facilitate the process of annotation, dedicated software could be used to produce meaning representations for tombstone images. Research has been carried out in this area of computational semantics for similar meaning representations (Wang & Xue, 2017; Van Noord & Bos, 2017), but these approaches assume textual input, not images of text. Parsing tombstone inscriptions raises new challenges, and here we give an overview of what we think could be potential showstoppers of this ambitious but exciting idea. An automatic tombstone parser has the potential to speed up human annotation, for instance in a setting where the parser produces a draft meaning representation, for correction by a human annotator.

³It could well be that ICONCLASS is better for this than WordNet. To give an example: ouroboros is not represented in WordNet, but it is present in ICONCLASS. On the other hand, a winged sandglass is not available in ICONCLASS but can be represented using a combination of two WordNet concepts.

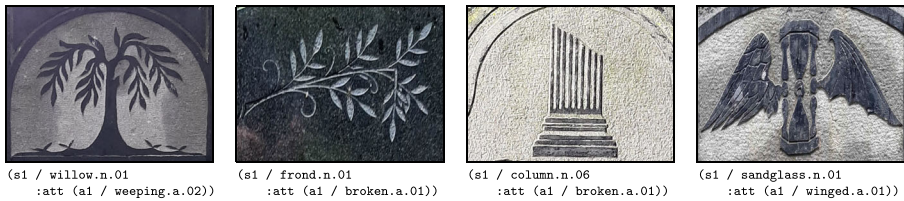


Fig. 8 Tombstone symbols with attributes, paired with their corresponding meaning representations

There are two extremes of natural language processing architectures that could be considered for this task. The first is a pipeline architecture, a sequence of tools dedicated to specific subtasks, typically where a component adds information to the output of the previous component in the pipeline (the first component would take the images as input; the last would output the meaning representation). The second is an architecture based on neural networks, where a machine learning model trained on pairs of input-output representations (here: images and their corresponding meaning representations) would predict a meaning representation for images not seen during the training stage. And, of course, anything in between these two extreme types of architectures is also possible. Here, we mainly consider the first type of architecture.

Considering a classic pipeline architecture, the first component faces the task of optical character recognition (OCR). There are various challenges here, as there is a tremendous variation in font type and size, background and foreground colour, orientation, and obscured or unreadable parts. It could also be challenging to map regular symbols, such as the star and cross to denote birth and death, to standard characters (see Fig. 6).

The next component would address the issue of segmentation: dividing characters into word tokens. This requires a proper interpretation of hyphenation (line breaks) and spaces: sometimes there is no spacing at all on small stones. Furthermore, parallel columns of texts on stones could introduce errors.

The next logical step would be to assign linguistic categories to the word tokens that are the result of the segmentation process. Recognising names of people, dates and locations (named entity recognition) is a first step, but this needs to be further extended for dealing with job descriptions, epitaphs, measurements and biblical references.

The interpretation component requires the normalisation of date expressions, the expansion of abbreviations (see Table 6), the resolution of toponyms (see Section 4.3), word sense disambiguation, pronoun resolution and coreference resolution. In other words, nearly all of the challenges of natural language understanding are present on tombstone inscriptions. On top of this, information is presented in a variety of languages, and code switching occurs as well.

Connections between entities are sometimes hard to establish because of long-distance dependencies. A case in point is a date of birth that does not belong to the closest person mentioned on the stone, as in the sequence “Here lies person A,

Table 6 Abbreviations found on tombstones in Groningen with their expansion and English translation

Abbreviation	Expansion	Translation
WED. / Wed. / WEDW	weduwe	widow
WEDN.	weduwenaar	widower
Geb. / GEB. / geb.	geboren	born
Overl. / OVERL. / overl. / Overl.den / OVL.	overleden	died
Gel. / GEL.	geliefd	beloved
GEM.	gemeente	municipality
ECHTG. / ECHTGEN. / Echtgen. / ECHTGE.	echtgeno(o)t(e)	spouse
GRON. / Gron.	Groningen	(city of) Groningen
D / D.l.	Duitsland	Germany
N.H.	Noord Holland	North Holland
O.	Oldambt	(region of) Oldambt
JOHs	Johannes	John

married to person B, born on date C". Without further context, "person B" could be linked erroneously to "date C".

Finally, coordinate structures may appear within dates, toponyms and personal names, where the coordinator is a parenthesis (Fig. 9). Coordination is a linguistic device that links together shared information, thereby reducing the amount of space required for the expression. Coordination is known as one of the hardest problems in natural language processing (Dahl & McCord, 1983), and the graphical coordination displayed on tombstones increases this difficulty even further.

If automatic approaches to tombstone parsing are developed, it is important to assess their quality. A standard way of doing this in natural language processing is to compare system output with "gold standard" representations; in other words, manually labelled meaning representations. For SPL, several tools have been developed to do this automatically, by computing precision and recall of matching triples (Cai & Knight, 2013). The harmonic mean of precision and recall gives a single score that measures the quality of the parser. A score of around 0.75, which is roughly what is achieved for text interpretation for similar meaning representations (Wang & Xue, 2017), would probably speed up the manual annotation of tombstone interpretation.

**Fig. 9** Examples of graphical coordination

6 Conclusion and future work

The information found on tombstones in the north of the Netherlands is diverse and stretches far beyond the name of the deceased and dates of birth and death. Commemorators, locations of birth and death (or other important places), family status, occupations, durations (someone's age or length of relationship or occupation) and references to biblical texts are common. Textual information is also regularly accompanied by symbolism, ranging from singular objects to complex collections of burial elements.

The expressive power required for a semantic formalism to capture the information found on tombstones in our dataset can be limited to a fragment of first-order logic without negation. The SPL formalism lends itself well for this purpose, as we have shown in this article.

A closed class of 32 two-place predicates is needed to establish fine-grained relations between the entities found on tombstones. As we have shown, WordNet provides a generic way of representing concepts, covering persons, locations, occupations, family relationships, measurements, organisations, biblical references and symbols. For our dataset of over 1,000 tombstones, a set of around 250 different concepts sufficed. The core formalism can be easily extended to existing third-party databases or classifications, including HISCO (Van Leeuwen et al., 2002), ICON-CLASS (Couprie, 1983) and GeoNames (Maltese & Farazi, 2013). This can be seen as a first step towards a standard for tombstone inscriptions for initiatives such as CIDOC-CRM (Biagetti, 2016) and Schema.org (Guha et al., 2016).

Not all phenomena have been integrated into our formalism. Information of cause of death, for example, is infrequent in our dataset, and more data is needed to design a generalised scheme compatible with existing classifications such as ICD-10 (Tobin et al., 2019). Another phenomenon that is not included is epitaphs (short phrases or texts in honour of the deceased person), as this requires further study (classification) and integration into SPL.

Computational approaches to interpreting tombstone inscriptions are likely to speed up the annotation process for datasets that are several orders of magnitude larger than the one presented in this work. They will need to deal with a large variety of complicated natural language processing issues, ranging from optical character reading and image recognition to segmentation, relation extraction, concept disambiguation, coreference resolution, and more. The dataset originating from this work will be released for research with this purpose in mind. Intelligent software that helps with annotation, and even imperfect approximations, could accelerate data labelling considerably.

There are various possible directions to take for further exploration. One exciting area is to consider tombstones of religions other than Christianity, such as gravestones of Jewish and Muslim communities. This is likely to require additional expertise to transcribe and interpret Hebrew and Arabic scripts. Another area is visualisation and consistency checking of the annotated data, to make the dataset 'AI-ready'. Exploring alternative ways of visualising the information of one or more, possibly related, tombstones, for instance by displaying them as timelines, will not only offer new perspectives on the data, but also provide means to capture contradictions, discover redundancies and detect anomalies.

Appendix A: Sources

Locations of cemeteries visited for the data collection of this study. Some cemeteries are anonymous, in which case the street name is provided.

No.	Cemetery Name	Town	Province
1	Kerkhoflaan	Emmen	Drenthe
2	Zuiderweg	Hoogeveen	Drenthe
3	Boerdijk	Nieuw-Amsterdam	Drenthe
4	Boterakkersweg	Sleen	Drenthe
5	Eswal	Vries	Drenthe
6	Kerkstraat	Beerta	Groningen
7	Sebastiaankerk	Bierum	Groningen
8	Engelberterweg	Engelbert	Groningen
9	Kerk van Heveskes	Farmsum	Groningen
10	Kerk Garmerwolde	Garmerwolde	Groningen
11	Esserveld	Groningen	Groningen
12	Zuiderbegraafplaats	Groningen	Groningen
13	Noorderbegraafplaats	Groningen	Groningen
14	Selwerderhof	Groningen	Groningen
15	Eshof	Haren	Groningen
16	Kerkstraat	Haren	Groningen
17	De Stille Hof	Hoogezand	Groningen
18	Middelberterweg	Middelbert	Groningen
19	Kerkhof Stefanuskerk	Noorddijk	Groningen
20	Noorderweg	Noordhorn	Groningen
21	Pollseweg	Noordlaren	Groningen
22	Ranumerweg	Obergum	Groningen
23	Oosternielandsterweg	Oosternieland	Groningen
24	Valgenweg	Oterdum	Groningen
25	Voorstraat	Oudeschans	Groningen
26	Kerklaan	Scheemda	Groningen
27	Stadsweg	Ten Boer	Groningen
28	Oldebertweg	Tolbert	Groningen
29	Leensterweg	Ulrum	Groningen
30	J.G. Pinksterstraat	Veendam	Groningen
31	Hoofdstraat	Vierhuizen	Groningen
32	Klokkentoren Westerdijkshorn	Westerdijkshorn	Groningen
33	Schilligeham	Winsum	Groningen
34	Klinckemalaan	Zuidhorn	Groningen
35	Franekerstraat	Bolsward	Friesland
36	Mariakerk	Buitenpost	Friesland
37	Noorder Begraafplaats	Leeuwarden	Friesland
38	Dominee H.W. Hundlingiuspad	Schiermonnikoog	Friesland
39	Kerkplein	Vlieland	Friesland
40	Kirchring	Bunde	Ostfriesland

Appendix B: Logic

The meaning representations that we use, Penman Interface Notation or Sentence Plan Language (SPL), can be simply translated to first-order logic (predicate logic). To do this, we first need to define the syntax of SPL:

Definition 1 (Syntax of SPL) $T ::= c \mid (x/P) \mid (x/P :R_1T_1 \dots :R_nT_n)$

Following this definition, SPL is recursively defined. An SPL representation can be a constant, a concept, or a concept related to one or more relations. The translation to first-order logic is defined recursively here, with the help of two cooperating translation functions. The first translation function, \mathcal{F} , maps an SPL representation into a formula of first-order logic (denoting a truth value); the second translation function, \mathcal{S} , maps an SPL representation into a formula of second-order logic (denoting a function from entities to truth values, written as a lambda term).

Definition 2 (Translation to First-Order Logic) $\mathcal{F}(c) = \exists x x=c$

$$\begin{aligned} \mathcal{F}((x/P)) &= \exists x P(x) \\ \mathcal{F}(x/P :R_1T_1 \dots :R_nT_n) &= \exists x (P(x) \wedge \mathcal{S}(T_1)(\lambda y.R_1(x,y)) \wedge \dots \wedge \mathcal{S}(T_n)(\lambda y.R_n(x,y))) \end{aligned}$$

Definition 3 (Translation to Second-Order Logic) $\mathcal{S}(c) = \lambda Q.Q(c)$

$$\begin{aligned} \mathcal{S}(x/P) &= \lambda Q.\exists x(P(x) \wedge Q(x)) \\ \mathcal{S}(x/P :R_1T_1 \dots :R_nT_n) &= \lambda Q.\exists x (P(x) \wedge \mathcal{S}(T_1)(\lambda y.R_1(x,y)) \wedge \dots \wedge \mathcal{S}(T_n)(\lambda y.R_n(x,y)) \wedge Q(x)) \end{aligned}$$

The translation can be illustrated using a simple example. Consider an SPL representation that states that there is a female person with the role of mother. Its translation proceeds as follows:

$$\begin{aligned} \mathcal{F}((p1 / female.n.01 :rol (r1 / mother.n.01))) &= \\ \exists p1(female.n.01(p1) \wedge \mathcal{S}((r1 / mother.n.01))(\lambda y.rol(p1,y))) &= \\ \exists p1(female.n.01(p1) \wedge \lambda Q.\exists r1(mother.n.01 \wedge Q(r1))(\lambda y.rol(p1,y))) &= \\ \exists p1(female.n.01(p1) \wedge \exists r1(mother.n.01 \wedge \lambda y.rol(p1,y)(r1))) &= \\ \exists p1(female.n.01(p1) \wedge \exists r1(mother.n.01 \wedge rol(p1,r1))) & \end{aligned}$$

Appendix C: Inference

Logic gives us the power to draw inferences in a controlled way. Inferences are useful for extending search capabilities or for performing systematic checks for inconsistencies in the data. Given the vocabulary of concepts and relations that we defined in our signature, we can make explicit what is implicit by defining logical axioms. In this appendix, we give some examples using standard first-order logic (using \exists for existential quantification, \forall for universal quantification, \wedge for conjunction, \vee for disjunction, \rightarrow for implication, \leftrightarrow for equivalence and \neg for negation).

We can make an explicit ontology of concepts by relating all concepts to each other, where the most general concepts (entity.n.01 and group.n.01) subsume more specific ones. For instance, we can state that male and female persons are persons, and that persons are entities, and that males are disjoint from females:

$$\begin{aligned} &\forall x((\text{male.n.02}(x) \vee \text{female.n.02}(x)) \rightarrow \text{person.n.01}(x)) \\ &\forall x(\text{person.n.01}(x) \rightarrow \text{entity.n.01}(x)) \\ &\forall x(\text{male.n.02}(x) \rightarrow \neg \text{female.n.02}(x)) \end{aligned}$$

We can apply a similar scheme to the roles in our inventory. We can state that wives and husbands are spouses, and that husbands are never wives:

$$\begin{aligned} &\forall x((\text{wife.n.01}(x) \vee \text{husband.n.01}(x)) \rightarrow \text{spouse.n.01}(x)) \\ &\forall x(\text{husband.n.01}(x) \rightarrow \neg \text{wife.n.01}(x)) \end{aligned}$$

We can add more constraints, and ensure that wives are always roles carried out by female persons, and husbands by males. And vice versa, we can also state that a female spouse is a wife, and a male spouse a husband.

$$\begin{aligned} &\forall x \forall y(\text{rol}(x,y) \wedge \text{wife.n.01}(y) \rightarrow \text{female.n.02}(x)) \\ &\forall x \forall y(\text{rol}(x,y) \wedge \text{husband.n.01}(y) \rightarrow \text{male.n.02}(x)) \\ &\forall x \forall y(\text{female.n.02}(x) \wedge \text{rol}(x,y) \wedge \text{spouse.n.01}(y) \rightarrow \text{wife.n.01}(y)) \\ &\forall x \forall y(\text{male.n.02}(x) \wedge \text{rol}(x,y) \wedge \text{spouse.n.01}(y) \rightarrow \text{husband.n.01}(y)) \end{aligned}$$

For the temporal relations, we can relate the *before* and *after* relations to each other, and define them as irreflexive and transitive:

$$\begin{aligned} &\forall x \forall y(\text{bef}(x,y) \leftrightarrow \text{aft}(y,x)) \\ &\forall x \forall y \forall z(\text{aft}(x,y) \wedge \text{aft}(y,z) \rightarrow \text{aft}(x,z)) \\ &\forall x \neg \text{aft}(x,x) \end{aligned}$$

More advanced axioms are possible that add further verification to the data. For instance, we can define an axiom that states that if someone is a widow of a person since a certain date, that person died on this date. Or, we can specify an axiom that expresses that every part of a composite image on a stone is also an image on that stone.

$$\begin{aligned} &\forall x \forall y \forall z(\text{widow.n.01}(x) \wedge \text{beg}(x,y) \wedge \text{tgt}(x,z) \rightarrow \text{dod}(z,y)) \\ &\forall x \forall y \forall z(\text{sym}(x,y) \wedge \text{prt}(y,z) \rightarrow \text{sym}(x,z)) \end{aligned}$$

There are two ways in which we can employ these axioms. Firstly, we can incorporate them as part of a search engine, to enable indirect searches (searching for a generic term would also yield more specific terms that entail it). Not all types of axioms will be suitable for this purpose. Secondly, axioms can be included as background knowledge for general purpose theorem provers and model builders for first-order logic. Theorem provers can be used to detect contradictions; model builders can be used to verify consistency and remove redundancies.

Appendix D: Groups

Collections of entities are represented by a special kind of entity: a group, with concept `group.n.01`. Groups have members, groups have a cardinality, and groups can contain sub-groups. At the top of the ontology, we distinguish between groups and non-groups, i.e., singular entities represented by `entity.n.01`.

$$\begin{aligned} &\forall x(\text{entity.n.01}(x) \rightarrow \neg \text{group.n.01}(x)) \\ &\forall x(\text{group.n.01}(x) \leftrightarrow \exists y \text{ qua}(x,y)) \\ &\forall x(\text{group.n.01}(x) \rightarrow \exists y(\text{entity.n.01}(y) \wedge \text{mem}(x,y))) \\ &\forall x\forall y(\text{sub}(x,y) \rightarrow \text{group.n.01}(x) \wedge \text{group.n.01}(y)) \\ &\forall x\forall y\forall z((\text{mem}(x,y) \wedge \text{sub}(y,z)) \rightarrow \text{mem}(x,z)) \end{aligned}$$

In our domain of tombstone inscriptions, we have two commonly used sets: people and measures. We consider people as groups of persons, and measures as groups of units of measurement.

$$\begin{aligned} &\forall x(\text{people.n.01}(x) \rightarrow \text{group.n.01}(x)) \\ &\forall x(\text{measure.n.02}(x) \rightarrow \text{group.n.01}(x)) \\ &\forall x\forall y(\text{people.n.01}(x) \wedge \text{mem}(x,y) \rightarrow \text{person.n.01}(y)) \\ &\forall x\forall y(\text{measure.n.02}(x) \wedge \text{mem}(x,y) \rightarrow \text{unit.n.01}(y)) \end{aligned}$$

Groups should not be confused with composite entities, even though this distinction is not always easy to make. Meronyms are modelled using the relation `:prt` (something has a part), but group membership is captured by `:mem`. A book with 20 chapters is represented as follows:

```
(b1 / book.n.01
  :prt (g1 / group.n.01
        :qua 20
        :mem (c1 / chapter.n.01)))
```

The tenth chapter of a book gets the following translation:

```
(c1 / chapter.n.01
  :ord 10
  :mem-of (g1 / group.n.01
           :prt-of (b1 / book.n.01)))
```

In the annotation, we use a shortcut, leaving out the membership relation, and stipulate that if x is the n th part of y , then x is a member of a group that is part of y . For instance, a reference to “Psalm 118:17–18” is represented as a group of two verses, both part of the same chapter.

```
:ref (g1 / group.n.01 :qua 2
      :mem (v1 / verse.n.01
            :ord 17
            :prt-of (c1 / chapter.n.01
                    :ord 118
                    :prt-of (b1 / Psalm.n.01)))
```

```

:mem (v2 / verse.n.01
      :ord 18
      :prt-of c1))

```

Groups of persons are also commonly expressed in tombstone inscriptions. An example is “our sister”, where the plural pronoun denotes a group of people. We represent this as:

```

(p1 / female.n.01
 :rol (r1 / sister.n.01
      :tgt (g1 / people.n.01)))

```

Here, an axiom is needed to trigger that if a role targets a group, then this target group has at least one member that fulfils this role. In other words, if Ann is a sister of Bea and Cai, then Ann is a sister of Bea, and Ann is a sister of Cai. The question is how strong to make this assumption: for at least one member of this group, or for all members? Consider the following, slightly more complicated example, “onze geliefde man en vader” (“our beloved husband and father”), which could be represented as follows:

```

(p1 / male.n.01
 :rol (r1 / husband.n.01
      :tgt (g1 / people.n.01))
 :rol (r2 / father.n.01
      :tgt g1))

```

Now, one might object to this representation, as it says that this male entity *p1* is the husband of a group of people *g1* and a father of the same group of people. If we make the group constraint weak (at least one member of the group plays the role), then this representation would be correct. If we make the group constraint strong (all members of the group play the indicated role), then we need to revise the representation as follows, where we include membership and sub-groups:

```

(p1 / male.n.01
 :rol (r1 / husband.n.01
      :tgt (p2 / female.n.01
           :mem-of (g1 / people.n.01)))
 :rol (r2 / father.n.01
      :tgt (g2 / group.n.01 :sub-of g1)))

```

Acknowledgements I would like to thank Bart Ramakers, Lidewijde de Jong, Tamara Dijkstra and Albert-Erik de Winter for valuable help in interpreting tombstone inscriptions. I am also grateful to Beatrice Alex, Clair Grover and Richard Tobin for providing information on the Digitising Scotland project, and to Tommaso Caselli for help with formulating SPARQL queries. Finally, I would like to thank the anonymous reviewers who gave valuable feedback on an earlier version of this article.

Data Availability The generated during the current research are available in the Tombreader project repository: <https://www.let.rug.nl/bos/tombreader/>

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahlers, D. (2013). Assessment of the accuracy of geonames gazetteer data. In *Proceedings of the 7th workshop on geographic information retrieval, GIR '13* (pp. 74–81). New York: Association for Computing Machinery.
- Antoniou, G., Groth, P., van Harmelen, F., & Hoekstra, R. (2012). *A semantic web primer. Information systems*, 3rd edn. Cambridge: MIT Press.
- Arenas, M., Gottlob, G., & Pieris, A. (2018). Expressive languages for querying the semantic web. *ACM Trans. Database Syst.*, 43(3).
- Baker, C. F., Fillmore, C. J., & Lowe, J.B. (1998). The Berkeley FrameNet project. In *36th annual meeting of the association for computational linguistics and 17th international conference on computational linguistics. Proceedings of the conference, Université de Montréal, Montreal, Quebec, Canada* (pp. 86–90).
- Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse, Sofia, Bulgaria* (pp. 178–186).
- Bateman, J. A. (1990). Finding translation equivalents: an application of grammatical metaphor. In *Proceedings of the 13th conference on computational linguistics* (pp. 13–18).
- Bateman, J. A., & Paris, C. L. (1989). Phrasing a text in terms the user understand. In *Proceedings of IJCAI* (pp. 1511–1517).
- Bateman, J. A., Kasper, R. T., Schuetz, J. F. L., & Steiner, E.H. (1989). A new view on the process of translation. In *Proceedings of the fourth conference on European chapter of the association for computational linguistics* (pp. 282–290).
- Biagetti, M. T. (2016). Un modello ontologico per l'integrazione delle informazioni del patrimonio culturale: Cidoc-crm. *Italian Journal of Library and Information Science*, 7(3), 43–77.
- Bond, F., & Foster, R. (2013). Linking and extending an open multilingual Wordnet. In *Proceedings of the 51st annual meeting of the association for computational linguistics (Volume 1: Long Papers)* (pp. 1352–1362). Sofia: Association for Computational Linguistics.
- Cai, S., & Knight, K. (2013). Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st annual meeting of the association for computational linguistics (Volume 2: Short Papers)* (pp. 748–752). Sofia: Association for Computational Linguistics.
- Carmack, S. D. (2002). *Your guide to cemetery research*. Cincinnati: Betterway Books.
- Couprie, L. (1983). Iconclass: an iconographic classification system. *Art Libraries Journal*, 8(2), 32–49.
- Dahl, V., & McCord, M. C. (1983). Treating coordination in logic grammars. *Computational Linguistics*, 9(2), 69–80.
- Ding, L., Zhou, L., Finin, T., & Joshi, A. (2005). How the semantic web is being used: An analysis of foaf documents. In *Proceedings of the 38th annual Hawaii international conference on system sciences* (pp. 1–10). <https://doi.org/10.1109/HICSS.2005.299>.
- Eckert, E. (1993). Language change: The testimony of czech tombstone inscriptions in praha, texas. In *Varieties of Czech: studies in Czech sociolinguistics*.
- In C. Fellbaum (Ed.) (1998). *WordNet. An electronic lexical database*. Cambridge: MIT Press.
- Gangemi, A., Guarino, N., Masolo, C., & Oltramari, A. (2003). Sweetening WORDNET with DOLCE. *AI Magazine*, 24(3), 13–23.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.) *Syntax and semantics: Vol. 3: speech acts* (pp. 41–58). San Diego: Academic Press.

- Guha, R., Brickley, D., & Macbeth, S. (2016). Schema.org: evolution of structured data on the web. *Communications of the ACM*, 59(2), 44–51.
- Heim, I. (1982). The semantics of definite and indefinite noun phrases. PhD thesis, University of Massachusetts.
- Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2006). OntoNotes: the 90% solution. In *Proceedings of the human language technology conference of the NAACL, companion volume: short papers, Stroudsburg, PA, USA* (pp. 57–60).
- Kamp, H. (1984). A theory of truth and semantic representation. In J. Groenendijk, T. M. Janssen, & M. Stokhof (Eds.) *Truth, interpretation and information, FORIS, Dordrecht ? Holland/Cinnaminson ? U.S.A.* (pp. 1–41).
- Kasper, R. T. (1989). A flexible interface for linking applications to penman's sentence generator. In *Proceedings of the DARPA speech and natural language workshop, Philadelphia* (pp. 153–158).
- Kew, T., Shaitarova, A., Meraner, I., Goldzycher, J., Clematide, S., & Volk, M. (2019). Geotagging a diachronic corpus of alpine texts: Comparing distinct approaches to toponym recognition. In *Proceedings of the workshop on language technology for digital historical archives, INCOMA Ltd., Varna, Bulgaria* (pp. 11–18).
- Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2008). A large-scale classification of English verbs. *Language Resources and Evaluation*, 42(1), 21–40.
- Krommenhoek, W. (2016). Graftsymboliek uit de negentiende en de eerste helft van de twintigste eeuw. De Biltse Grift.
- Leidner, J. L. (2008). *Toponym resolution in text : annotation, evaluation and applications of spatial grounding of place names*. Boca Raton: Universal Press.
- Long, C. E. (2016). Disambiguating the departed. using the genealogist's tools to uniquely identify the long dead and little known. *LRTS*, 60(4), 236–247.
- Maltese, V., & Farazi, F. (2013). A semantic schema for geonames.
- Matthias, G. F. (1967). Weathering rates of portland arkose tombstones. *Journal of Geological Education*, 15(4), 140–144.
- Palmer, M., Kingsbury, P., & Gildea, D. (2005). The proposition bank: an annotated corpus of semantic roles. *Computational Linguistics*, 31(1), 71–106.
- Saller, R., & Shaw, B. (1984). Tombstones and roman family relations in the principate: civilians, soldiers and slaves. *Journal of Roman Studies*, 74, 124–156.
- Slenk, H. (1969). Jan utenhove's psalms in the low countries. *Nederlands Archief voor Kerkgeschiedenis / Dutch Review of Church History*, 49(2), 155–168.
- Streiter, O., & Goudin, Y. (2011). Bury your past, shovel it under: Histories and caterpillars on taiwan's graveyards. In *First workshop on documenting and researching gravesites in Taiwan* (pp. 27–48).
- Streiter, O., Voltmer, L., & Goudin, Y. (2007). From Tombstones to corpora: Tsm1 for research on language, culture, identity and gender differences. In *Proceedings of the 21st Pacific Asia conference on language, information and computation (PACLIC), Seoul National University, Seoul, Korea* (pp. 450–458).
- Streiter, O., Lin, A. M., Yen, S. K. R., Hsu, E., Wang, Y., & Goudin, Y. (2008). Place names on Taiwan's tombstones: facts, figures theories. In *Workshop on the relationship between the distribution of languages or dialects in Taiwan and racial migration*.
- Streiter, O., Goudin, Y., & Huang, C.J. (2010a). Thakbong, digitalizing taiwan's tombstones for teaching, research and documentation. In *TELDAP 2010 - The Int. Conf. on Taiwan e-learning and digital archives program* (pp. 146–157).
- Streiter, O., Huang, C. J., & Goudin, Y. (2010b). Places of shared histories: spatial patterns of placename types on Taiwan's tombstones, GIS-IDEAS 2010, Hanoi, Vietnam.
- Tobin, R., Farrow, E., Grover, C., & Alex, B. (2019). Automatic coding of occupation and cause-of-death records. *International Journal of Population Data Science*, 4(3).
- Toscana, M. (2019). Crowdsourcing Irish History. In *III national AIPH conference, Università degli Studi della Campania*.
- Van Assem, M., Gangemi, A., & Schreiber, G. (2006). RDF/OWL representation of WordNet. <http://www.w3.org/TR/wordnet-rdf/>, W3C Public Working Draft.
- Van Hage, W. R., Malaisé, V., Segers, R., Hollink, L., & Schreiber, G. (2011). Design and use of the simple event model (SEM). *Journal of Web Semantics*, 9(2), 128–136. <https://doi.org/10.1016/j.websem.2011.03.003>. <http://www.sciencedirect.com/science/article/pii/S1570826811000199>, provenance in the Semantic Web.

- Van Leeuwen, M. H., Maas, I. M., & Miles, A. (2002). *HISCO. Historical international standard classification of occupations*. Leuven: Leuven University Press.
- Van Noord, R., & Bos, J. (2017). Neural semantic parsing by character-based translation: Experiments with abstract meaning representations. *Computational Linguistics in the Netherlands Journal*, 7, 93–108.
- Veit, R. F., & Nonestied, M. (2008). *New Jersey cemeteries and tombstones: history in the landscape*. New Brunswick: Rutgers University Press.
- Wang, C., & Xue, N. (2017). Getting the most out of AMR parsing. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 1257–1268). Copenhagen: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D17-1129>. <https://www.aclweb.org/anthology/D17-1129>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.