From Tombstones to Corpora: TSML for Research on Language, Culture, Identity and Gender Differences

Oliver Streiter[a], Leonhard Voltmer[b], Yoann Goudin[c]

[a]National University of Kaohsiung, Taiwan, ostreiter@nuk.edu.tw
[b]European Academy Bolzano/Bozen, Italy, lvoltmer@eurac.edu
[c]Ecole des Hautes Etudes en Sciences Sociales, France, yoanngoudin@yahoo.com.fr

Abstract. Tombstone inscriptions represent a genre which yields insights into cultures and languages. Applying the idea of linguistic corpora to tombstones, we propose to create tombstone corpora as sustainable resource for the study of languages and cultures. For the annotation of tombstone corpora, we propose TSML, the Tombstone-Markup-Language, developed during the annotation of tombstones from Taiwan plus, in addition, some from China, Indonesia and Europe. We develop and discuss our conceptual framework in the annotation of tombstones with its cultural, linguistic and psychological perspectives. We will outline possible research strategies which can be followed with TSML-annotated corpora, ranging from the analysis of word meanings, to models of identity and the comparison of cultures with respect to the patterns of reference systems they provide.

Keywords: Tombstones, corpora, XML, TSML, Tombstone-Markup-Language, Taiwan.

Tombstone inscriptions *represent a genre which, due to the moment they represent, provide profound insights into culture and language.* When the trivia of life don't matter anymore and the cullets of life are swept together in a few strokes in marble, the tomb is frequently the only agent and the only trace in a battle between conflicting identities, social relations, conceptual systems, mythologies and religions.

Plate 1 to 4: Tombstones in Taiwan. From left to right, a sinicized aborigine tomb, a Japanese-style Han tomb, a christianized Han tomb and a de-sinicized aborigine tomb.

Tombstones *can be found worldwide.* Their form and inscription follow ethnic and religious

traditions, cf. Rath 1986, Frembgen 1998, The Hindu 2005, Benninghausen 2005. Notwithstanding global trends, tombstones preserve local customs. Even on a small island like Taiwan, tombstones in the North and South, East and West are different, blending the flavors of ethnic or religious traditions with local craftsmanship (cf. Clark 1989). Traditions in contact borrow from each other and create particular forms as shown in Plates 1 to 4.

*Creating a **corpus of tombstones**, as opposed to other research designs based on tombstones, requires most investments, but is also quite promising.* First, a corpus can reveal facts, such as local, temporal, ethnic, religious, social or gender-related differences, that cannot be learned from individual tombstones. In addition, a corpus, when properly balanced, paves the way to innumerable investigations beyond the initial research questions for which the corpus had been designed. Third, a corpus with photos or videos as integral part can be continuously annotated, opening new perspectives with each new annotation. Finally, corpora from different resources can be used in comparative studies.

*Tombstone corpora however cannot bridge the gap between the **rapid extinction of cultures and languages** (Wurm 1991) and the lack of activities in documenting them.* Although one might hope that tombstones will still be recoverable after a culture had died, factors like urbanization, industrialization, tourism or acid rain threaten the existence of these mute witnesses. With a handful of tombstones found after 100 years, any systematic comparison across regions, ethnicities or time periods will be difficult. Our study on Taiwanese tombstones confirms the precarious state of tombstones older than 60 years.

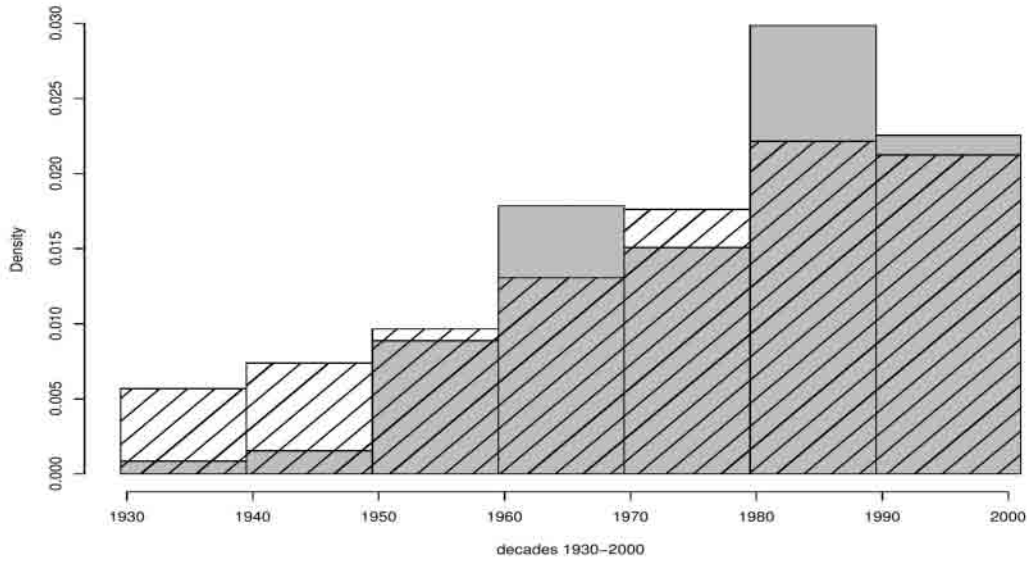

Figure 1: The comparison of the density of tombstones (grey) and of the population (shaded) shows the loss of older tombstones in Taiwan. Data based on 3000 tombstones.

Different grass-root activities have sprung up, e.g. in the US and Australia, to *preserve the **cultural heritage** of tombstones by photographing and transcribing them* (e.g. http://www.rootsweb.com/~cemetery). However, the nature of the transcriptions determines the use one can make of them. Unstructured transcriptions, for example, leave too much ambiguity for automatic analysis. The word 'Brown' might be a name or a color, 'Miller' a name or profession. 'Brown' thus should be annotated as 'name' and 'Miller' as 'profession'. However,

much more is involved when annotating tombstones, especially if on strives for referential annotations, as opposed to textual or editorial annotations. Referential annotations are crucial for cross-cultural and cross-linguistic comparisons, as only the references have to be merged or mapped without touching the corpus annotation as such.

***XML** (Bray et al. 2004) is, without question, the best supported meta-language for corpus annotation* and with XCES (CES 2000) and TEI (Sperberg-McQueen & Burnard 2002) two widely used corpus annotation languages are available. To our knowledge, however, no XML language for tombstone corpora has been developed. There is the EpiDoc project however, that aims at the annotation of epigraphs (Anderson et al 2007) and develops a rich scheme for the textual elements on the basis of TEI. However, EpiDoc does not yet provide an annotation framework beyond the text, such as the description of graves, graveyards or their ethnic and religious environments. In addition, EpiDoc stresses the individuality of the object, given the function of epitaphs as revelation of the individual personality, cf. Edgette 1989.

A corpus, however, serves different purposes. *In a **corpus** an individual stone or any individual feature is meaningless.* When annotating a corpus, we annotate only those features which, beyond the purposes of data management and data retrieval, enter a system of meaningful oppositions. Non-meaningful oppositions like the temperature of the tombstone, which create connotations (cf. Fages 1968) are ignored. In terms of statistics, a feature is not annotated as long as there is no conjecture of a correlation with another feature.

*Simplicity, uniformity and flexibility of TSML is achieved by using the ** element** in combination with a **type-attribute*** as shown in Figure 2. We do not specify any constraints on the hierarchy of *type*-attributes. There exist tombs without graveyards, graveyards without tombs, tombstones without tombs, tombs without tombstones, tombs within tombs and a tombstone-side, for example as photo, without the stone. Symbols, photos and texts can be at all levels.

```
<tsml>
 <div type='graveyard' north='55.34254' east='13.55456'religion='christianism'>
   <media mime_type='image/jpg' src='http://....../>
   <div type='church' background_color='green' set_up='1954-10-05' />
   <div type='graveyard_section' ethnicity='ami people'>
    <div type='tomb' direction='180' background_color='red' orientation='downhill'
         set_up='1962-09-02'>
      <div type='tomb_side' side='inside' vertical='90' direction='270>
        <div type='image' description='fish'/>
      </div>
      <div type='tombstone'>
       <div type='version' set_up='1962-09-02' status='lost'/>
       <div type='version' set_up='2002-01-05'>
         <media mime_type='image/jpg' src='http://....../>
         <div type='tombstone_side' background_color='white' foreground_color='black'
              writing_direction='t2br2l' script='han-zi'>
          <div type='photo' floating='top' size='6cm' description='male'>
            <media mime_type='image/jpg' src='http://....../>
          </div>
          <div type='symbol' floating='top' size='10'
               description='presbitarian cross'/>
          <div type='text' floating='top'>
           <div type='text' floating='right'>
            <div type='p'>text goes here</div>
           </div>
           <div type='text' floating='right' language='ami' script='katakana'>
            <div type='p'>text goes here</div>
```

Figure 2: The basic XML-structure of TSML based on *** *** elements and ***type***-attributes, describing here an imaginary tomb.

Table 1: Values of the *type*-attribute of the <*div*> element in TSML.

	Explanation
graveyard	Site where tombs are located.
graveyard_section	Graveyard sections may relate to different ethnicities or religions.
church, temple, ...	A building related to cults which contains graves or is located in a graveyard.
tomb	A site containing the remains of one or more deceased.
tomb_side	An inner or outer wall of a tomb.
tombstone	The tombstone as 3-dimensional object.
tombstone_side	A 2-dimensional view on the tombstone.
tombstone_unit	Relatively independent units within a tombstone or a tombstone-side.
text, p, w, c, stroke	Containing mainly text.
image	Containing mainly an image.
symbol	Containing mainly a non-figurative symbol.
photo	Containing mainly a photo.

For all <*div*> elements, the attributes listed in Table 2 are assumed to be inherited (to cascade) from the mother <*div*> to the daughter <*div*> in the absence of the attribute at the level of the daughter. Some of these values are derived external sources (GPS, compass, map, archives).

Table 2: Attributes of the <*div*> element to be cascaded. Attributes marked '*' have been suggested in Debartolo Carmack 2002.

	Examples	Explanation
name*	Taipei Fude	Graveyards, graves may have official or unofficial names.
description		Free text input.
location*	Taipei	Name of town, city, township where the entity is located.
caretaker*	Taipeishi	
caretaker address*		Caretakers might be contacted for additional information.
composition*	marble	Basic material: *marble, slate, granite, sandstone, limestone, metal, brick, concrete, ceramics*.
status*	abandoned	Useful to explain data loss, data endangerment. Values: *abandoned, maintained, overgrown, eroded, broken, lost*.
north	5.88789	Latitude as decimal WGS84 datum (cf. NIMA 97).
east	52.87465	Longitude as decimal WGS84 datum (cf. NIMA 97).
elevation	417	The elevation above mean see level in meter.
direction	90	Cardinal direction: *0=360*=North, *90*=East, *180*=South, ...
orientation	downhill	Non-compass directional system: *uphill, downhill, upcoast, downcoast, upstre am, downstream, landward, seaward, lakeward, mountainward, streetward, concentric*.
side	inside	*Inside, outside* with respect to the outer border of an object.
vertical	90	*90*=vertical, *0*=horizontal, wall and roof respectively.
set-up	2001-09-01	Time of construction/ building/ writing/ photographing.
floating	right	Relative position within the mother <*div*> element, observer position opposite to the orientation, as in CSS absolute position (Bos et al. 2007). Alternative values: *right, left, top, bottom*.
display	block	Display similar to CSS (Bos et al. 2007). Values: *block, inline, list-item, superimposed, none.*
background- color	red	The color of the background, as in Çelik and Lilley 2003.
foreground-color	red, green, ...	The color of the foreground, as in Çelik and Lilley 2003.
religion	Buddhism, ...	The main religious orientation according to XNLRDF[1].
ethnicity	Hakka, Ami	The main ethnicity according to XNLRDF.
language	eng, deu	ISO 639-3 language codes, cf. XNLRDF.
writing-direction	t2bl2r, l2rt2b t2br2l, ..	Top-to-bottom left-to-right (Chinese), right-to-left top-to-bottom (English), top-to-bottom left-to-right (Mongolian), cf. XNLRDF.
script	Latin, Arabic	The set of characters or signs used according to XNLRDF.

A corpus should be balanced (Biber et al. 1998). This requirement corresponds to, for example, correct sampling methods in surveys and assures that t*he findings in the corpus or survey can be* ***generalized*** *to the overall population one is interested in.* Although for a tombstone corpus, criteria of balancedness might be better to define than for a text corpus, e.g. collecting two photo per 10000 tombs, it is impossible to achieve balancedness through sampling in the field. It is impossible to know all graveyards or all tombs and those tombs that one might find, may be inaccessible or decayed. In addition, naïve balancedness is not what we are out for. We want a

[1] XNLRDF, the Natural Language Research Description Framework, cf. Streiter et Stuflesser 2006.

tombstone corpus to have different granularities for different subsets of the data. We want to photograph and annotate graveyards of minorities exhaustively. In relatively uniform Han-communities of major cities we content us to take samples. *If we would follow uniform sampling conditions for the entire corpus, no comparison inside smaller groups would be statistically possible. We therefore use weights to achieve a* ***numerical balancedness***.

A ***weight*** *stipulates how many items of the population are represented by one sampled item.* Using the given, estimated or interpolated values of $inhabitants_{ly}$ and $life\text{-}expectancy_{ly}$ for locality *l,* year *y* and graveyard *g,* (1) estimates the population of tombstones. The sample-size for a location is derived from the sample-size of its graveyards, where graveyards might belong to more than one location (2). The margin is the size of the population of this location not already claimed by its sub-locations (3). The weight of a tombstone for its directly associated location calculates as (4). The weight for larger geographic or administrative region ("*How many tombstone does this tombstone represent in South-Taiwan*?") is estimated on the basis of the weights of the location, the relative population of the location and the region and how much of the margin is claimed by the region itself (4). Although this model is based on estimations, *a* ***calculated balancedness based on census data*** *with the possibility to have different granularities in different sub-corpora allows for a much better control of the sampling than with sampling criteria for text corpora.* Actually, this model actively guides our sampling as it points us to locations, regions or time periods relatively underrepresented.

(1) $population_{l,y} = inhabitants_{l,y} / life\text{-}expectancy_{l,y}$

(2) $sample\text{-}size_{l,y,g} = \sum^{g}(sample\text{-}size_{y,g} / \sum locality_{g})$

(3) $margin_{region,y} = population_{region,y} - \sum^{location} population_{location,y}$

(4) $weight_{location,y,g} = margin_{location,y} / sample\text{-}size_{location,y,g}$

(5$weight_{region,y,g} = weight_{location,y,g} \times ((population_{region,y} - (weight_{region,y,g2} \times sample\text{-}size_{region,y,g2})) / \sum^{l} population_{l,y})$.

Putting these formulas to work, we could estimate the population of *Kunshen* as 5000. With an estimated life-expectancy of 50 in 1980 we would expect to find 100 tombs of 1980 (1). Actually we found 20, but as the graveyard is shared with *Xishu* the sample size for *Kunshen* is 10 (2). As *Kunshen* has no sub-locations with a proper graveyard, the margin is 10 (3). The weight for the tombstones of *Kunshen* for *Kunshen* for 1980 is thus 10, i.e. one tombstone found represents 10 tombstones of the population. *Kunshen* is part of Tainan City, while *Xishu* is not. Tainan city has graveyards at local levels (*Anping* and *Kunshen)* plus graveyards at the city level. We assume that from the 700000 inhabitants of Tainan 100000 are related to *Anping* and *Kunshen*. The margin of Tainan City is thus 600000 inhabitants with a life expectancy of 60 = 10000. The population of Tainan City would then be 700000/60=11666. As the population of *Anping* and *Kunshen* is 100000/50 = 2000, the weight of the tombstones of *Kunshen* for *Tainan City* for 1980 is thus 10 x (11666/2000) = 58. The calculation continues for the next administrative or geographical level.

```
<div type='graveyard' name='kunshen/xishu'>
  <weights>
    <weight year='1980' loc='kunshen' value='10'/>
     <weight year='1980' loc='tainan city' value='58'/>
```

Figure 3: Weights assigned to tombstones with respect to the location and the year in TSML.

Like any historic document, tombstones may become ***unreadable***. If something is totally unreadable, this has to be marked as '*unreadable*' as opposed to not yet annotated. If something is partially readable, say like the given name *Deb?rah,* where *?* stands for the unreadable, we almost for certain recognize the name, but we would falsify the data if we would annotate that we have read *Deborah.* It would be also inappropriate to encode the entire name as unreadable, as such a data loss cannot be made up, e.g. for a foreign language like Hebrew: דְּ?וֹרָה. In TSML, keeping track of what has been written and, if not perfectly readable, what the interpretations of the remaining traces are, is the first level of description. Interpretations can be refined with probabilities.

As ***readability****, especially of Chinese characters or Egyptians hieroglyphs varies below the character level*, i.e. at the level of radicals or strokes, we cannot use character indices as done in the annotation of the American National Corpus (cf. Ide & Romary 2006). If, for example, we can read a vertical stroke in a position where we expect a Chinese number, we might interpret this as 一 (1) or, if an additional scratch represents a vertical stroke, as 十 (10). We describe such phenomena using the <*analysis*> and <*interpretation*> elements. The <*analysis*> element gives a syntagmatic analysis of the mother <*div*> while repeating its content. The repetition of the content together with the *display='superimpose'* allows to refer to a grapheme below the character level. Those elements marked as *include='yes'* fall under the analysis, as described by the attributes of the <*analysis*> element, in Figure 4 the markers of a 3rd person singular subject. The <*analysis*> element can be repeated for multiple annotations of the same <*div*>.

```
<div type='s'>He eats fish
 <analysis type='3ps'>
  <div type='w' include='yes'>He</div>
  <div type='c' include='no'> </div>
  <div type='verb_root' include='no'>eat</div>
  <div type='infl_morph' include='yes'>s</div>
  <div type='c'> </div>
  <div type='w'>fish</div>
 </analysis>
</div>
```

Figure 4: An example of a discontinuous structure annotated using the *analysis*-element.

The ***annotation of alternative solutions*** *is quite standard in XML*, cf. CES 2000 or Good & Hendryx-Parker 2006. In TSML, the <interpretation> element lists possible paradigmatic choices, where each choice is a possible interpretation of the perceived traces. Preferences in their selection can be marked by *selected='yes', selected='no'* or *probability='0.8'*.

```
<div type='p' ref_id='1' ref_type='year' ref_system='ROC calendar'>民國五
  <div type='num'>
  <interpretation>
   <div type='num' selected='yes'>一
    <analysis type='num'>
     <div type='stroke' include='yes' >一</div>
    </analysis>
   </div>
   <div type='num' language='' probability='' selected='no'> 十
    <analysis type='num'>
      <div type='stroke' include='yes' >一</div>
     <div type='stroke' include='yes' display='superimpose'> | </div>
    </analysis>
   </div>
  </interpretation>...
```

Figure 5: An example of how different interpretations (一 or 十) are derived from different analyses of the traces on the tombstone. The entire <*div*> is then assigned a reference.

The content of the <*div*> elements may have ***references***. Different <*div*> elements may have the same *reference*, as it is the case for the name of a person in two languages, cf. Plates 5-7. In such a case, language A and language B are ***reference systems***, i.e. systems that allow to calculate the ***references***. The annotation and analysis of reference systems will lead directly into the heart of languages, cultures and societies as we shall show below.

The references*, like times, persons and places are entities which in past, present or future lead an imagined or real existence independently from the grave or the tombstone.* In TSML, we store them therefore in a small conceptual structure. Relations among references are references themselves. Temporal references, for example, are obtained by translating the date we find in a specific calendar in the corpus into the date of a calendar of reference within the conceptual structure. In the same way we can map the names of a city in the corpus onto a reference system of imagined or real cities, or personal names to a reference repository of persons. The interest in annotating references derives from the historical, geographical or sociological facts they reflect. Such facts might be apprehended through tombstones or they provide background information for the interpretation of other data on the tombstones.

Plate 5-7: Two languages (Hebrew and German), two scripts (Hebrew and Gothic), two writing directions (r2lt2b and l2rt2b) and one name for one person in **(5)**. Two languages (Japanese and Mandarin), two scripts (Katakana and Hanzi) and two names for one person in **(6)**. Two languages (German and Italian), one script, two character types and two names (Italian calque created during Italian fascism, German original name added after fascism) for one person.

We identify ***reference systems*** *with social mediators which shape experience and awareness for members of a culture.* For a description of mediation as psychological process see, among others, Wertsch 1988. Thus, for some analyses of tombstones, neither the exact date (*reference)*, nor the exact wording of a date might be interesting, but instead what kind of calendar (*reference system)* is used, since a calendar is such a social construct which mediates psychological processes. This means that people use this construct to think, when calculating dates or time spans, or use this construct to organize and control their behavior.

Similarly, one might analyze whether symbols are taken from a Christian symbol repository or from a Jewish symbol repository, instead of analyzing what the symbol means (the *reference).* The way that people are referred to, how their gender is marked, the language, the script, the writing direction, all are additional, reference systems that merit an analysis. As shown above for the names, reference systems do not cooccur randomly. Instead they cluster into cultural groups. The clusters can be derived automatically from a corpus, provided that the reference systems have been annotated.

```
<tsml>
 <tsml_conceptual_structure>
  <person id='1' ref_system='family_given_latin' args='0'
   value='Brown Bill'/>
  <person id='2' ref_system='family_given_latin' args='0'
   value='Brown Mary'/>
  <father_of id='1' ref_system='father_child' args='2'
   arg1_type='person' arg2_type='person' arg1_id='1', arg2_id='2' />
  <day id='1' ref_system='gregorian' args='0' value='2007-12-27'/>
  <year id='1' ref_system='gregorian' args='0' value='1962'/>...
```

Figure 6: References, reference types (person, father, day, year) and standard reference systems in the TSML conceptual structure, where the annotations point to.

Another, important category for annotation and analysis are ***reference types****, intentional abstractions of references, as for example, 'person', 'father_of', 'day', 'day of birth_of'.* Thus, while 'Bill, father of Mary' has a *reference* (an entry in the TSML conceptual structure as shown in Figure 6) within a *reference system* (one of the definition of fatherhood in a society), the reference type, marked in the corpus as * * points to a class of elements of the ontological component of TSML across cultures and languages.

Reference types are annotated if we want to analyze and compare the meaningful elements of a tombstone (symbols, words, expressions, arrangement) across cultures and languages. We thus annotate all meaningful elements which identify the nature of a maybe larger textual element as pointing to the *<father_of>* elements in the conceptual structure with the reference type *ref-type=father_of*'. This way we can retrieve and analyze all meaningful expressions of a given reference type in all languages and all modes (image, text, symbol) in their context.[2] In addition to linguistic expressions and symbols, colors or spatial arrangements may have references and

[2]Note, that this model requires all quasi-synonymous expressions of interest for a linguist to be defined in relation to *references* within TSML. This might be too strong a claim and imply that one has to invoke the apparatus of the reference system during annotation, even if the reference system allows for one reference only. We can tweak this by making the *reference system* and *reference* optional. If there is, for example, only one honorific reference, something like 'honored' without any gradiant, we can write ** .**

thus can be linked to *reference types*. During the annotation process the color of a segment *** *** might be elaborated into *** ,*** according to the meaning of the color. If referring structures are discontinuous, we can use the <*analysis*> element together with the referential attributes to identify those elements which share these referential attributes. If the string *father of'* that identifies the reference type (*father_of*) is a substring of the string '*Bill, father of Mary',* giving the reference, the referential attributes can be annotated at different levels, identifying the linguistic elements responsible for the reference type, the reference system or the reference. For the interpretation of theses separated features, we can percolate them up to the smallest common upper bound. Thus referential features do not cascade.

In our research on Taiwanese tombstones, still another reference type, that of the ***local origin*** is of central importance. *Many Taiwanese families actually have the possibility to chose between different reference systems, the* ***Tanghao****, a mythological place name in Northern China, a place name in South China from where the ancestors immigrated (****Jiguan****) or a place name in Taiwan (****Taiwan diming****).* The name might not be important for an analysis, the reference system of the origin however hints to identities that communities maintain. In addition, we expect this reference system (Tanghao/Jiguan/Taiwan diming) to cluster with other reference systems, such as the calendar (Japanese/Chinese/Republican/Gregorian). As with names, we find combinations of reference systems, e.g. Tanghao with Jiguan or combinations of different calendars on one stone. The outcome of this research on correlations will be subject to further publications.

Plate 8-10: Left to right. Three reference systems for the reference type 'origin', the Tanghao, the Jiguan and the place name in Taiwan. Each reference system is a system of paradigmatic choices, possibly incompatible with those of the other system.

Research on other reference types such as ***gender-markers***, as in the following examples in Plates 15-17 from the churchyard of the village Gries, now a suburb of the city of Bolzano in the Italian Alps, will extend the analysis into other aspects of a culture. Research steps involve (a) the recognition or reconstruction of the referens system with its paradigmatic values, (b) the establishment of the relationship, conflict, incommensurability between alternative reference systems of the same referenence type and (c) *to cluster automatically over the corpus reference systems of different referens types to develop the pattern of the branching structure of related referens types as finger print of a culture.* Different cultures might show similar patterns, e.g. two languages, two writing systems, one assocatiated with ethnicity or religion, the other with economical or political systems, or different systems.

Plate 11-14: Left to right, top to bottom. Four calendars (reference systems) found on Taiwanese tombstones: The Japanese calendar, the traditional Chinese calendar, the Republican calendar and the Gregorian calendar. All possible combinations of calendars can be found on tombstones, particularly popular is the combination of Chinese and Republican calendar. As preliminary data show, however, specific calendars correlate to locations, ethnicities and religion.

Plate 15: Transcription: "*Here rests the virgin Anna Hilpold, died 19 March 1917 in her 68th year of life. R.I.P.*" The gender marker "*virgin*" highlights that despite the age, a woman was referred to through a 'male' reference system. This can be the absence of any male, like in Plate 15, a male-like occupation like "*Master"*, as in **Plate 16**: "*Here rests in God Mrs Barbara Schlögele born Wiedenhofer Master in Gries ...*", or the husband as in **Plate 17:** "*Here rest the married couple Sebastian Unterkofler*", where "*Sebastian"* is the husband's name.

In this paper we presented our concerns and motivations in constructing an annotation language for tombs and tombstones. Calling this annotation ***concept-based*** is partially correct only as we do not necessarily construct one concept hierarchy. *If reference systems are not commensurable, we might end up in incompatible conceptual fragments.* Our approach is, at the same time ***profoundly linguistic***. *However, contrary to classical structuralist approaches we have a cultural layer between the language and the paradigmatic choices of the language. This cultural layer determines the reference system and with it, which system of paradigmatic choices is made available.* Our initial claim to research in language and culture is thus no empty slogan. *According to our model,* ***culture*** *is epitomized in more often than not a cluster of reference systems.* ***Language*** *is one of the physical bodies of culture, providing material in the*

form of paradigmatic choices distinct from those of another reference system. At the same time, reference systems are cognitive structures, especially if they map onto incommensurable conceptual structures. The ***identity*** *of a person and a society is affected by the way clustered reference systems branch,* whether branching points are close or far from fundamentally personal reference systems. ***Societies*** *finally show similarities with respect to the patterns of clusters they form,* where the reference types allow us to map and align different reference systems across cultures and languages.

References

Benninghausen, R. 2005. Friedho fe als Quellen fu r Fragen des Kulturwandels: Grabkultur von Yeziden und Aleviten in Deutschland mit Seitenblick auf die Tu rkei. In: R. Langer, R. Motika, M. Ursinus (eds): *Migration und Ritualtransfer: Religio se Praxis der Aleviten, Jesiden und Nusairier zwischen Vorderem Orient und Westeuropa*, Peter Lang, Frankfurt a M. pg. 247–288.

Biber, D., S. Conrad, R. Reppen. 1998. *Corpus Linguistics. Investigating Language Structure and Use.* Cambridge University Press.

Bos, B., T. Çelik, I. Hickson, H.W. Lie. 2007. Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. W3C. http://www.w3.org/TR/CSS21/cover.html, accessed 24.9.07.

Bray, T., J. Paoli, C.M. Sperberg-McQueen, E. Maler, J. Cowan, F. Yergeau (eds.). 2004. *Extensible Markup Language (XML) 1.1*, W3C. http://www.w3.org/TR/xml11, accessed 22.6.07.

CES, Corpus Encoding Standard, Part 5: Encoding Linguistic Annotations. 2000. http://www.cs.vassar.edu/CES/CES1-5.html, accessed 22.6.07.

Çelik, T. & C. Lilley. 2003. CSS3 Color Module, W3C Candidate Recommendation 14 May 2003. http://www.w3.org/TR/css3-color, accessed 24.9.07.

Clark, E.W. 1989/1992. The Bigham Carvers of the Carolina Piedmont: Stone Images of an Emerging Sense of American Identity. In: R.E. Meyer (ed.), *Cemeteries & Gravemarkers, Voices of American* Culture. Utah State University Press.

Debartolo Carmack, S. 2002. *Your Guide to Cementary Research*, Betterway Books.

Edgette, J.J. 1989/1992. The Epitaph as Personality Revelation. In: R.E. Meyer (ed.), *Cemeteries & Gravemarkers, Voices of American* Culture. Utah State University Press.

Elliott, T. (ed). 2007. EpiDoc: Guidelines for Structured Markup of Epigraphic Texts in TEI. http://www.stoa.org/epidoc/gl/5, accessed 22.6.07.

Fages, J.B. 1968. *Comprendre le structuralisme*. Privat, Toulouse.

Frembgeni, J.W. 1989. Religious Folk Art as an Expression of Identity: Muslim Tombstones in the Gangar Mountains of Pakistan. In: *Muqarnas*, Vol. 15, pp. 200-210.

Good, J. & C. Hendryx-Parker. 2006. Modeling Contested Categorization in Linguistic Databases. In: *2006 EMELD Workshop on Digital Language Documentation, Tools and Standards, the State of the Art*. Michigan State University in East Lansing, Michigan, June 20-22.

Ide, N., L. Romary. 2006. Representing Linguistic Corpora and Their Annotations. Proceedings

of the *Fifth Language Resources and Evaluation Conference (LREC)*, Genoa, Italy.

The Hindu. Buddhist Tombstone discovered at Chittayam, *The Hindu, online edition of India's National Newspaper* 8.7.2005, http://www.hindu.com/2005/07/08/stories/2005070811950300.htm, accessed 23.6.07.

NIMA Technical Report TR8350.2, "Department of Defense World Geodetic System 1984, Its Definition and Relationships With Local Geodetic Systems", 3rd Edition, 4 July 1997. http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html, accessed 24.9.07.

Rath, G. 1986. Hebrew Tombstone Inscriptions and Dates. In: *Chronicles (Newsletter of the Jewish Genealogical Society of Philadelphia)*, Vol. 5, No. 1 (Spring 1986), pages 1-4.

Sperberg-McQueen, C.M., L. Burnard (eds). 2002 *Guidelines for Text Encoding and Interchange.* University of Oxford.

Streiter, O., M. Stuflesser. 2006. Design Features for the Collection and Distribution of Basic NLP-Resources for the World's Writing Systems. In: *Intl. Workshop Towards a Research Infrastructure for Language Resources*, LREC Workshop, Genova, Italy, 22 May 2006.

Wertsch, J.V. 1988. *Vygotsky and the Social Formation of Mind*. Harvard University Press.

Wurm, S. (ed). 2001. *Atlas of the World's Languages in Danger of Disappearing*. Paris, UNESCO.