The Association for Computers and the Humanities (ACH)

The Association for Computational Linguistics (ACL)

The Association for Literary and Linguistic Computing (ALLC)

Guidelines for Electronic Text Encoding and Interchange

Edited by

C. M. Sperberg-McQueen and Lou Burnard

TEI P2, Chapter 19 CE
Certainty and Responsibility
Text Encoding Initiative
Chicago, Oxford

September 17, 1993

Note

This is the current draft (as of September 17, 1993) of one section of the *Guidelines for Electronic Text Encoding and Interchange*, version 2 (also known as TEI P2), produced by the Text Encoding Initiative. This fascicle contains:

- a table of contents for TEI P2, indicating the status of each section as of the release date of this fascicle
- a draft for one section of TEI P2
- entries derived from this section for the reference list of tags of TEI P2 (if any)
- a User Response and Comment Form

Please feel free to send us your comments on the contents of this draft and their applicability to your needs. Comments may be sent to the TEI-L list server for general discussion, or directly to either of the TEI editors (adresses on the User Response and Comment form). Bear in mind in reading, however, that TEI P2 is a reference manual, not a tutorial in the use of electronic text, and unavoidably assumes either a certain amount of prior knowledge or patience with unfamiliar technical material. Note also that some matters which may appear obscure in one section may well be treated in more detail elsewhere in the full text of TEI P2.

TEI P2, like P1, is a working draft. The observant will find several typographic and presentational flaws in the current text: for example, cross-references are not properly resolved in the LaTeX versions and some parts of the reference material are not yet complete. We hope to rectify as many as possible of these flaws in the coming months, but have decided not to allow them to delay any further the release of the draft. We hope that those interested in the Guidelines will find the current state of the text readable and usable enough for the purpose of substantive comment. As far as possible, all comments received on this section will be taken into account in the preparation of the next version of the Guidelines, which is due by the end of 1993.

Updates on the status of the draft as a whole will be distributed automatically to all subscribers of the Listserv list TEI-L. To subscribe to TEI-L, send electronic mail to the address LISTSERV@UICVM (or Listserv@uicvm.uic.edu) containing the single line

```
subscribe tei-l J. Smith
```

(substituting your name for “J. Smith”). TEI-L is also the appropriate place to pose questions or offer public comments on the TEI guidelines and other relevant issues. The TEI-L file server contains all the sections of TEI P2 thus far released, as well as other TEI materials. Its contents are shadowed at various other sites around the world; for more information consult one of the following documents:

- TEI ED J8, “Obtaining the Second Version of the TEI Guidelines (TEI P2)” (describes how to retrieve electronic copies of TEI P2 and the various formats they are available in)
- TEI ED J9, “Obtaining Paper Copies of the Second Version of the TEI Guidelines (TEI P2)” (describes how to request paper copies of TEI P2, for those without electronic mail access)

These documents are available from the TEI-L file server or through the editors at the addresses given on the User Response and Comment Form.

Overall Table of Contents for TEI P2

The following list shows the expected contents of TEI P2 as a whole and indicates the current status of each chapter.

Contents

Part 4

Additional Tag Sets

Chapter 19

CE: Indications of Certainty and Responsibility

Encoders of text often find it useful to indicate that some aspects of the encoded text are problematic or uncertain, and to indicate who is responsible for various aspects of the markup of the electronic text. These Guidelines provide two methods of recording uncertainty about the text or its markup:

- the **<note>** element defined in section 6.8, “Notes, Annotation, and Indexing,” [in separate fascicle]may be used, with a value of “certainty” for its *type* attribute.
- the **<certainty>** element defined in this chapter may be used to record the nature and degree of the uncertainty in a more structured way.

There are three methods of indicating responsibility for different aspects of the electronic text:

- the TEI header records who is responsible for an electronic text by means of the **<resp>** element and other more specific elements (**<author>**, **<sponsor>**, **<funder>**, **<principal>**, etc.) used within the **<titleStmt>**, **<editionStmt>**, and **<revisionDesc>** elements.
- the **<note>** element may be used with a value of “resp” or “responsibility” in its *type* attribute.
- the **<respons>** element defined in this chapter may be used to record fine-grained structured information about responsibility for individual tags in the text.

To use the **<note>** and **<resp>** elements, no special steps are needed, since they are defined in the core tag set and header respectively. To use the **<certainty>** and **<respons>** elements, the additional tag set for certainty and responsibility must be selected; this is done by defining the parameter entity *tei.certainty* with the value “INCLUDE”, as shown in the example below:

```
<!DOCTYPE tei.2 system 'tei2.dtd' [
   <!ENTITY % tei.certainty 'INCLUDE' >
   <!-- other modifications to the DTD here ... -->
]>
 ...
```

19.1 Levels of Certainty

Many types of uncertainty may be distinguished. The **<certainty>** element is designed to encode the following sorts:

- a given tag may or may not correctly apply (e.g. a given word may be a personal name, or perhaps not)
- the precise point at which an element begins, or ends, is uncertain
- the value to be given for an attribute is uncertain

- content supplied by the encoder (such as the expansion of an abbreviation marked by the **<abbrev>** tag) is uncertain.
- the transcription of a source text is uncertain, perhaps because it is hard to read or hard to hear. (This sort of uncertainty is also handled by the **<unclear>** element in section 20.11, "Damage, Illegibility, and Supplied Text," [in separate fascicle]).

The following types of uncertainty are *not* indicated with the **<certainty>** element.

- a number or date is imprecise
- the text is ambiguous, so a given passage has several possible interpretations
- a transcriber, editor, or author wishes to indicate a level of confidence in a factual assertion made in the text
- an author is not sure if the sentence she has chosen to start a paragraph is really the one she wants to retain in the final version

Precision of numbers and dates is handled in section 6.4, "Names, Numbers, Dates, Abbreviations, and Addresses," [in separate fascicle] and chapter 22, "Additional Tags for Names and Dates," [in separate fascicle]; well-defined ambiguity is handled with alternations in feature-structure values in chapter 18, "Additional Tag Set for Feature Structures," [in separate fascicle]. Uncertainty about the truth of assertions in the text and other sorts of authorial and editorial uncertainty about whether the content is satisfactory are not handled by the **<certainty>** element at this time (though they may be expressed using **<note>**).

19.1.1 Using Notes to Record Uncertainty

The simplest way of recording uncertainty about markup is to attach a note to the element or location about which one is unsure. In the following (invented) paragraph, for example, an encoder might be uncertain whether to mark "Essex" as a place name or a personal name, since both might be plausible in the given context:

> Elizabeth went to Essex. She had always liked Essex.

Using **<note>**, the uncertainty here may be recorded quite simply:

```
<persName>Elizabeth</> went to
<placeName>Essex</>.
She had always liked <placeName>Essex</>.
<note type='uncertainty' resp='MSM'>It is not
clear here whether <mentioned>Essex</mentioned>
refers to the place or to the nobleman. -MSM</note>
```

Using the normal mechanisms, the note may be associated unambiguously with specific elements of the text, thus:

```
<persName>Elizabeth</> went to
<placeName id=p1>Essex</>.
She had always liked <placeName id=p2>Essex</>.
<note type='uncertainty' resp='MSM' target='p1 p2'>
It is not clear here whether <mentioned>Essex</mentioned>
refers to the place or to the nobleman.  If the latter,
it should be tagged as a personal name. -MSM</note>
```

The advantage of this technique is its relative simplicity. Its disadvantage is that the nature and degree of uncertainty are not conveyed in any systematic way and thus are not susceptible to any sort of automatic processing.

19.1.2 Structured Indications of Uncertainty

To record uncertainty in a more structured way, susceptible of at least simple automatic processing, the **<certainty>** element may be used:

<certainty> : indicates the degree of certainty or uncertainty associated with some aspect of the text markup. Attributes include:
- *target* : points at the elements whose markup is uncertain.
- *locus* : indicates the precise location of the uncertainty in the markup: applicability of the element, precise position of the start- or end-tag, value of a specific attribute, etc. Suggested values include:
  - *#gi* : uncertain whether the element used actually applies to the passage.
  - *#startloc* : start-tag may not be correctly located.
  - *#endloc* : end-tag may not be correctly located.
  - *#location* : both the start-tag and the end-tag may not be correctly located.
  - *[attribute name]* : the value given for the attribute named is uncertain.
  - *#transcribedContent* : the content of the element may not be a correct transcription of the source text.
  - *#suppliedContent* : the content of the element may not have been correctly supplied by the reader, e.g. as in the cases of **<corr>** and **<abbrev>** elements.
- *given* : indicates conditions assumed in the assignment of a degree of confidence.
- *degree* : indicates the degree of confidence assigned to the aspect of the markup named by the *locus* attribute.
- *assertedValue* : provides an alternative value for the aspect of the markup in question—an alternative generic identifier, transcription, or attribute value, or the ID of an **<anchor>** element (to indicate an alternative starting or ending location). If a n *assertedValue* is given, the confidence level specified by *degree* applies to the alternative markup specified by *assertedValue*; if none is given, it applies to the markup in the text.
- *desc* : further describes the uncertainty in prose, perhaps indicating its nature, cause, or the justification for the degree of confidence asserted.

The **<certainty>** element may be used to record doubts about the proper encoding of “Essex” in several ways of varying precision. To record merely that we are not certain that “Essex” is in fact a place name, as it is tagged, we use the *target* attribute to identify the element in question, and the *locus* attribute to indicate what aspect of the markup we are uncertain about (here: whether we have used the correct element type):

```
Elizabeth went to <placeName id=p1>Essex</>.

<!-- ... elsewhere in the document ... -->
<certainty target=p1 locus='#gi' desc='possibly not a placename'>
```

Because it is linked to the location of the uncertainty by an IDREF, the **<certainty>** element will typically be included in the same SGML document as its target. It may be placed adjacent to the target element, or elsewhere in the document.

To record the further information that we estimate, subjectively, that there is a 60% chance of "Essex" being a place name here, we can add a value for our *degree* of confidence (usually a number between 0 and 1, representing the estimated probability):

```
Elizabeth went to <placeName id=p1>Essex</>.

<!-- ... -->
<certainty target=p1 locus='#gi'
           degree='0.6'
           desc='possibly not a placename'>
```

To record the estimate of one expert (AR) that there is a 60% chance of "Essex" being a place name here, and a 40% chance of its being a personal name, we use the *resp* attribute to associate the certainty information with AR. To specify the second probability, we use a second **<certainty>** element, pointing at the same location in the text, in which an alternative choice of generic identifier (here: "persName") is given as the value of the *assertedValue* attribute:

```
Elizabeth went to <placeName id=p1>Essex</>.

<!-- ... -->
<certainty target=p1 locus='#gi'
           degree='0.6'
           desc='probably a placename, but possibly not'>
<certainty target=p1 locus='#gi'
           degree='0.4'
           assertedValue='persName'
           desc='may refer to the Earl of Essex'>
```

Finally, we may wish to make our probability estimates contingent upon some condition. In the passage "Elizabeth went to Essex; she had always liked Essex," for example, we may feel there is a 60% chance that the county is meant, and a 40% chance that the earl is meant. But the two occurrences of the word are not independent: there is (we may feel) no chance at all that one occurrence refers to the county and one to the earl. We can express this by using the *given* attribute to list the SGML identifiers of **<certainty>** elements.

```
Elizabeth went to <placeName id=p1>Essex</>.
She had always liked <placeName id=p2>Essex</>.

<!-- ... -->
<!-- 60% chance that P1 is a placename,
     40% chance a personal name. -->
<certainty id=cert-1
           target=p1 locus='#gi' degree='0.6'
           desc='probably a placename, but possibly not'>
<certainty id='cert-2'
           target=p1 locus='#gi' degree='0.4'
           assertedValue='persName'
           desc='may refer to the Earl of Essex'>

<!-- 60% chance that P2 is a placename,
     40% chance a personal name.
     100% chance that it agrees with P1. -->
<certainty given='cert-1'
           target=p2 locus='#gi' degree='1.0'
           desc='if P1 is a placename, P2 certainly is'>
<certainty given='cert-2'
           target=p2 locus='#gi' degree='1.0'
           assertedValue='persName'
           desc='if p1 refers to the Earl of Essex, so does P2'>
```

When *given* conditions are listed, the **<certainty>** element is interpreted as claiming a given degree of confidence in a particular markup given the assertional content of the **<certainty>** elements indicated—that is, *if the markup described in the indicated **<certainty>** elements is correct.*

Conditional confidence may be less that 100%: given the sentence “Ernest went to old Saybrook”, we may interpret “Saybrook” as a personal name or a place name, assigning a 60% probability to the former. If it is a place name, there may be a 50% chance that the place name actually in question is “Old Saybrook” rather than “Saybrook”, while if it is correctly tagged as a personal name, it is much more likely (say, 90% certain) that the name is “Saybrook”. This state of affairs can be expressed using the **<certainty>** element thus:

```
Earnest went to <anchor id=a1> old
<persName id=p1>Saybrook</persName>.

<certainty id=c1 target=p1 locus='#gi' degree='0.6'>
<certainty given=c1 target=p1 locus=startloc degree='0.9'>

<certainty id=c2 target=p1 locus='#gi'
           assertedValue='persName' degree='0.4'>
<certainty given=c2 target=p1 locus=startloc degree='0.5'>
<certainty id=c3
           given=c1 target=p1 locus=startloc
           assertedValue='a1' degree='0.5' >
```

In this case, the *assertedValue* on **<certainty>** element c3 is an IDREF to an **<anchor>** element at the alternate starting point for the element.

Multiplying the numeric values out, this markup may be interpreted as assigning specific subjective probabilities to three different ways of marking up the sentence:

```
Earnest went to old <persName>Saybrook</>.   (0.6 * 0.9, or 0.54)
Earnest went to old <placeName>Saybrook</>.  (0.4 * 0.5, or 0.20)
Earnest went to <placeName>old Saybrook</>.  (0.4 * 0.5, or 0.20)
```

The probabilities do not add up to 1.00 because the markup indicates that if “Saybrook” is (part of) a personal name, there is a 10% likelihood that the element should start somewhere other than the place indicated, without however giving an alternative location; there is thus a 6% chance (0.1 * 0.6) that none of the alternatives given is correct.

If an attribute value is uncertain, the *locus* attribute takes as its value the name of the attribute in question. In this example, there is only a 50% chance that the question was spoken by participant A:

```
<u id=u1 who=A>Have you heard the election results?</u>
<!-- ... -->
<certainty target=u1 locus=who degree='0.5'>
```

Doubts about whether the transcription is correct may be expressed by assigning to *locus* the value “#transcribedContent”. For example, if the source is hard to read and so the transcription is uncertain:

```
I have a <emph id=p1>gub</emph>.
<certainty target=p1 locus='#transcribedContent' degree='0.5'>
```

Degrees of confidence in the proper expansion of abbreviations may also be expressed, by using the value “#suppliedContent”:

```
You will want to use <expan id=e1 abbr='SGML'>Standard
Generalized Markup Language</expan> ...

<!-- ... -->
<certainty target=e1 locus='#suppliedContent' degree='0.9'>
```

The *assertedValue* attribute should be used to provide an alternative value for whatever aspect of the markup is in doubt: an alternative generic identifier, or the ID of an alternative starting or ending point, as already shown, an alternative attribute value, or alternative element content, as in this example:

```
I have a <emph id=p1>gub</emph>.

<certainty target=p1 locus='#transcribedContent'
           assertedValue='gun'
           degree='0.8'
           desc='a gun makes more sense in a holdup'>
```

Since attribute values have no internal substructure, the *assertedValue* attribute is useful for specifying alternative transcriptions only in relatively restricted circumstances (specifically, when the alternate reading has no elements nested within it). More robust methods of handling uncertainties of transcription are the **<unclear>** element and the **<app>** and **<rdg>** elements described in chapter 21, “Additional Tag Set for Critical Apparatus,” [in separate fascicle].

There are many subtleties in the indication of uncertainty and the **<certainty>** element does not provide a fully general solution, or even an unproblematic one. In particular no claim is made that this mechanism can express every sort of uncertainty that an encoder may wish to assert, nor that it is wholly free from ambiguities. However, the **<certainty>** element does allow for indications of uncertainty to be structured with at least as much detail and clarity as appears to be currently required in most ongoing text projects. Encoders who wish to disambiguate some use of this element, or to extend the recommended attribute values, may document these assumptions or special vocabulary. It is expected that in the future more adequate systems for expressing uncertainty will be developed. These may extend the **<certainty>** element or they may make use of the feature-structure encoding mechanisms described in chapter 18, “Additional Tag Set for Feature Structures,” [in separate fascicle].

The **<certainty>** element and the other TEI mechanisms for indicating uncertainty provide a range of methods of graduated complexity. Simple expressions of uncertainty may be made by using the **<note>** element. This is simple and convenient, and can accommodate either discursive unstructured indication of uncertainty, or complex structured project-specific expressions of uncertainty. In general, however, unless special steps are taken, the **<note>** element does not provide as much expressive power as the **<certainty>** element, and in cases where highly structured certainty information must be given, it is recommended that the **<certainty>** element be used.

As noted, the **<certainty>** element may be used for simple unqualified indications of uncertainty, in which case only the *locus* and **<target>** might be specified. The other attributes may be used to provide fuller information. They may take any string of characters as value, but it is recommended that the recommended values be used wherever possible; if they are not appropriate in a given situation, encoders are urged to provide their own controlled vocabulary and document it in the **<encodingDesc>** or **<tagUsage>** elements of the TEI header.

The **<certainty>** element has the following formal declaration:

```
<!-- 19.1.2:  Certainty and uncertainty                      -->
<!-- Text Encoding Initiative: Guidelines for Electronic      -->
<!-- Text Encoding and Interchange. DRAFT Version 2. 1992-93. -->

<!-- Copyright (c) 1990, 1992, 1993 ACH, ACL, ALLC.           -->
<!-- Permission to copy in any form is granted for use with   -->
<!-- TEI-aware systems and applications, provided this        -->
<!-- notice is included in all copies.                        -->

<!-- These materials may not be altered; modifications to     -->
<!-- these DTDs should be performed as specified in chapter   -->
<!-- MD of the Guidelines.                                    -->
```

```
<!-- These materials subject to revision. Current versions    -->
<!-- available from the Text Encoding Initiative.             -->


<!ELEMENT certainty     - O  EMPTY                              >
<!ATTLIST certainty          %a.global;
          target             IDREFS              #REQUIRED
          locus              CDATA               #REQUIRED
          assertedValue      CDATA               #IMPLIED
          desc               CDATA               #IMPLIED
          given              CDATA               #IMPLIED
          degree             CDATA               #IMPLIED       >
<!-- ... declarations from section 19.2                        -->
<!--     (Responsibility for markup)                           -->
<!--     go here ...                                           -->
```

19.2 Attribution of Responsibility

In general, responsibility for the transcription and markup of an electronic text is borne by those named in the header: the title statement, the edition statement(s), and the revision history all allow (by means the **<resp>** element) for the specification, in general terms, of who is responsible for a given aspect of the electronic text.

In some cases, however, more detailed element-by-element information may be desired, in order to distinguish, for example, between the individuals responsible for transcribing the content and those responsible for determining that a given word or phrase constitutes a proper noun. Where such fine-grained attribution of responsibility is required, the **<respons>** element may be used:

<respons> : identifies the individual(s) responsible for some aspect of the markup of some particular element(s). Attributes include:
- *target* : gives the SGML identiifer(s) of the element(s) for which some aspect of the responsibility is being assigned.
- *locus* : indicates the specific aspect of the markup for which responsibility is being assigned. Suggested values include:
  - *#gi* : responsibility for the claim that the element is of the type indicated by the markup
  - *#location* : responsibility for the claim that the element begins and ends where indicated
  - *#startloc* : responsibility for the claim that the element begins where indicated
  - *#endloc* : responsibility for the claim that the element ends where indicated
  - *[attribute name]* : responsibility for the claim that the attribute named has the value given in the markup
  - *#transcribedContent* : responsibility for the transcription of the element content
  - *#suppliedContent* : responsibility for the contents supplied by the encoder (corrections, expansions of abbreviations, etc.)
- *resp* : identifies the individual or agency responsible for the indicated aspect of the electronic text.

desc : gives a brief prose note supplying any additional information which should be recorded

This element allows one or more aspects of the markup to be attributed to a given individual. The *target* and *locus* attributes function as they do on the **<certainty>** element described in section 19.1, “Levels of Certainty,” on page 3: the *target* attribute points at a particular SGML element (or set of elements), and *locus* indicates the particular aspect of the encoding of those elements, for which responsibility is to be assigned. The suggested values may be combined as appropriate: to indicate, for example, that RC is responsible for transcribing an illegible word, and that AR is responsible for identifying that word as a proper noun, the text might be encoded thus:

```
Earnest went to old <persName id=p1>Saybrook</persName>.
<!-- ... -->
<respons target=p1 locus='#transcribedContent' resp=RC>
<respons target=p1 locus='#gi #location' resp=AR>
```

Many elements bear specialized *resp* or *agent* attributes, which have specific meaning which varies from element to element; the **<respons>** element should be reserved for the general aspects of responsibility common to all text transcription and SGML markup, and should not be confused with the more specific attributes on individual elements.

The formal declaration of the **<respon>** element is this:

```
<!-- 19.2:  Responsibility for markup                         -->
<!ELEMENT respons       - O  EMPTY                              >
<!ATTLIST respons            %a.global;
          target             IDREFS             #REQUIRED
          locus              CDATA              #REQUIRED
          resp               CDATA              #REQUIRED
          desc               CDATA              #IMPLIED        >
<!-- This fragment is used in sec. 19.1.2                      -->
```

Part 7

Alphabetic Reference List of Tags

Class Name: **analysis**

Description: defines additional analytic attributes specifying alternate structural arrangements for the components of a text.

Attributes:

equiv (i.e. equivalent): specifies the identifiers of elements considered structurally equivalent to the one bearing the attribute.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

exclude: specifies the identifiers of elements that are considered excluded from the structure when the element bearing this attribute is selected.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

select: specifies which of several mutually exclusive elements actually occurs in a structure.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks: I wish I understood what this was for.
This attribute is optional.

Part: additional tag set for analysis and interpretation.
Member of classes:
DTD file:

Discussed in 17, “Additional Tag Set for Simple Analysis,” [in separate fascicle].

<certainty>

Description: indicates the degree of certainty or uncertainty associated with some aspect of the text markup.

Attributes:

target: points at the elements whose markup is uncertain.
Data type: IDREFS
Value: one or more SGML ID values.
Default value: #REQUIRED
Example:

```
Elizabeth went to <persName id=p1>Essex</persName>
<certainty target=p1 locus=gi degree='0.6'>
```

Remarks: If more than one IDREF is given, the **<certainty>** element is interpreted as applying to all. If no ID is present on the element being annotated, the attribute should give the ID of a **<ptr>** element which points at the element being annotated; for further discussion of this indirect pointing mechanism, see chapter “REFID= not defined” target=SA>.
This attribute is required.

locus: indicates the precise location of the uncertainty in the markup: applicability of the element, precise position of the start- or end-tag, value of a specific attribute, etc.
Data type: CDATA
Suggested values include:
#gi : uncertain whether the element used actually applies to the passage.
#startloc : start-tag may not be correctly located.
#endloc : end-tag may not be correctly located.
#location : both the start-tag and the end-tag may not be correctly located.
[attribute name] : the value given for the attribute named is uncertain.
#transcribedContent : the content of the element may not be a correct transcription of the source text.
#suppliedContent : the content of the element may not have been correctly supplied by the reader, e.g. as in the cases of **<corr>** and **<abbrev>** elements.
Default value: #REQUIRED
Example:

Remarks: The “#” distinguishes the terms of the controlled vocabulary from possible collisions with attribute names. Extensions to this vocabulary should also use this prefix.
This attribute is required.

assertedValue: provides an alternative value for the aspect of the markup in question—an alternative generic identifier, transcription, or attribute value, or the ID of an **<anchor>** element (to indicate an alternative starting or ending location). If an *assertedValue* is given, the confidence level specified by *degree* applies to the alternative markup specified by *assertedValue*; if none is given, it applies to the markup in the text.
Data type: CDATA
Value: generic identifier, attribute value, location (e.g. indicated by an IDREF to an **<anchor>** element or to an **<xptr>** element), or other appropriate alternative value.
Default value: #IMPLIED
Example:

```
Elizabeth went to <persName id=p1>Essex</persName>
<certainty target=p1 locus='#gi'
           assertedValue='place' degree='0.2'>
```

Remarks: This attribute makes it possible to indicate the degree of confidence in a specific alternative to some aspect of the markup. In the example above the encoder is expressing the likelihood (.2) that the generic identifier should be **<place>** rather than **<persName>**, which is the coded element.
This attribute is recommended.

desc: further describes the uncertainty in prose, perhaps indicating its nature, cause, or the justification for the degree of confidence asserted.
Data type: CDATA
Value: a prose description of how and why the markup is uncertain.
Default value: #IMPLIED
Example:

```
Elizabeth went to <persName id=p1>Essex</persName>
<certainty target=p1 locus='#gi'  degree='0.2'
desc='As this was written in the Spring
it is probably the Earl, not the town'>
```

Remarks: In a given project, it may be possible to enumerate a finite list of recognized types and causes of uncertainty; in such cases, it will be useful to control the vocabulary used in this attribute, to aid later mechanical manipulation. It is not possible to suggest such a controlled vocabulary for general use.
This attribute is recommended.

given: indicates conditions assumed in the assignment of a degree of confidence.
Data type: CDATA
Value: a characterization of the conditions which are assumed in the assignment of a degree of confidence. This may be in prose.
Default value: #IMPLIED
Example:

```
<!-- in the header, hand 'h1' is identified
     as that of 'msm' ... -->
<hand id=h1 scribe=msm>

<!-- in the text, the scribe has corrected 'Wessex'
     into 'Essex' ... -->
Elizabeth went to <corr id=c1 sic='Wessex' resp=msm>Essex</corr>.

<!-- we are 60% certain that hand h1 is 'msm',
     and 90% certain that if h1 is 'msm', then
     it is 'msm' who corrected 'Wessex'
     into 'Essex'.  -->
<certainty id=h1 target=h1 locus=scribe degree='0.6'>
<certainty target=p1 locus=resp given=h1 degree='0.9'>
```

Remarks: A project may wish to control the vocabulary used in this attribute.

The envisioned typical value of this attribute would be an IDREF of another **<certainty>** element or list (conjunction) of such IDREFs. It may thus be possible to construct probability networks by chaining **<certainty>** elements together. Such networks would ultimately be grounded in unconditional **<certainty>** elements (with no value for *given*. The semantics of this chaining would be understood in this way: if a **<certainty>** element is specified, via an IDREF, as the assumption, then it is not the attribution of uncertainty that is the assumption, but rather the assertion itself. For instance, in the example above, the first **<certainty>** element indicates that the confidence in the identification of the new scribe as “msm”. The second indicates the degree of confidence that “Essex” is a personal name, given that the new scribe is “msm”. (N.B. the given in the second **<certainty>** element is not the assertion that the likelihood that msm is the new scribe is 0.6, but simply the assertion that msm is the new scribe; this is a recommended convention to facilitate building networks.

The ambitious encoder may wish to attempt complex networks or probability assertions,.experimenting with references to other elements or prose assertions, and deploying feature structure connectives such as **<alt>**, **<join>**, and **<not>**. However we do not believe that the **<certainty>** element gives, at this time, a comprehensive ambiguity-free system for indicating certainty.
This attribute is recommended.

degree: indicates the degree of confidence assigned to the aspect of the markup named by the *locus* attribute.
Data type: CDATA
Value: Values of *degree* might be yes or no, the reals between 0 and 1, or traditional characterizations such as “doubtful”, “circa”, etc. Generally we recommend decimal numbers between 0 and 1, with the larger the number the greater the degree of confidence in the assertions; 0 representing “certainly false” and 1 representing “certainly true”.
Default value: #IMPLIED
Example:

This attribute is optional.

Example:

Part: additional tag set for.
Member of classes:
DTD file:
Data description: Empty.
May occur within:
May contain: EMPTY
Declarations:

```
<!ELEMENT certainty     - O  EMPTY                         >
<!ATTLIST certainty          %a.global;
          target             IDREFS           #REQUIRED
          locus              CDATA            #REQUIRED
          assertedValue      CDATA            #IMPLIED
          desc               CDATA            #IMPLIED
          given              CDATA            #IMPLIED
          degree             CDATA            #IMPLIED       >
```

Discussed in 19, “CE: Indications of Certainty and Responsibility,” on page 3.
This tag is optional.

Class Name: **certainty**
Description: defines additional attributes which are enabled for all elements when the additional tag set for certainty and uncertainty is selected.
Attributes: [None: global and inherited attributes only.]
Part: additional tag set for.
Member of classes:
DTD file:
Discussed in 19, “CE: Indications of Certainty and Responsibility,” on page 3.

Class Name: **fs** (i.e. feature-structures)
Description: defines additional attributes for all elements, which are enabled when the additional tag set for feature-structure notation is selected.
Attributes: [None: global and inherited attributes only.]
Remarks: [No attributes declared yet.]
Part: additional tag set for analysis and interpretation.
Member of classes:
DTD file:
Discussed in 18, “Additional Tag Set for Feature Structures,” [in separate fascicle].

Class Name: **global**
Description: defines attributes common to all elements in the TEI encoding scheme.
Attributes:

id (i.e. identifier): provides a unique identifier for the element bearing the ID value.
Data type: ID
Value: any valid SGML name (in the reference concrete syntax, a name must begin with a letter and contain only letters, numeric digits, hyphen, and full stop).
Default value: #IMPLIED
Example:

```
<p id=names>Paragraph with the ID <mentioned>names</>....
<p id=dates>Paragraph with the ID <mentioned>dates</>....
```

Remarks: The *id* attribute may be used to specify a canonical reference for an element; see section 6.9, “Reference Systems,” [in separate fascicle].
This attribute is optional.

n (i.e. number, name, etc.): gives a number (or other label) for an element, which is not necessarily unique within the document.
Data type: CDATA
Value: any string of characters; often, but not necessarily, numeric.
Default value: #IMPLIED
Example:

Remarks: The *n* attribute may be used to specify the numbering of chapters, sections, list items, etc.; it may also be used in the specification of a standard reference system for the text.
This attribute is optional.

lang (i.e. language): indicates the language of the element content, usually using a two- or three-letter code from ISO 639.
Data type: IDREF

Value: The value must be the identifier specified for a writing system declaration declared in the TEI header, as described in section "REFID= not defined" target=HD>.
Default value: %INHERITED
Example:

```
<p lang=EN>The only surviving work by <person>Ari</>
(died 1148) is the ten-page <title
lang=IS>&Iacute;slendingab&oacute;k</> (<title
lang=LA>Libellus Islandorum</>), written in the early
twelfth century.
```

Remarks: If no value is specified for *lang*, the *lang* value for the immediately enclosing element is inherited; for this reason, a value should always be specified on the outermost element (**<tei.2>**).
This attribute is optional.

rend (i.e. rendition or presentation): indicates how the element in question was rendered or presented in the source text.
Data type: CDATA
Value: any string of characters; if the typographic rendition of a text is to be systematically recorded, a systematic set of values for the *rend* attribute should be defined.
Default value: #IMPLIED
Example:

Remarks: These Guidelines make no binding recommendations for the values of the *rend* attribute; the characteristics of visual presentation vary too much from text to text and the decision to record or ignore individual characteristics varies too much from project to project. Some potentially useful conventions are noted from time to time at appropriate points in the Guidelines.
This attribute is optional.

Remarks: The global attributes described here are made part of the attribute definition list declaration of each element by including the string "%a.global" in each such declaration. Some global attributes are made available when certain base or additional tag sets are selected; these are incorporated into the global attributes by references to the appropriate parameter entities. When the tag sets in question have not been selected, the parameter entities in question expand to the empty string.
Part: base tag set for common core features.
Member of classes: clterm cllink clana clfs clcert clspoken
DTD file:
Discussed in 3.5, "Global Attributes," [in separate fascicle].

Class Name: # linking

Description: defines additional attributes for hypertext and other linking, which are enabled for all elements when the additional tag set for linking is selected.
Attributes:

corresp (i.e. correspondents): specifies the identifiers of corresponding elements.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

link (i.e. link): specifies the identifiers of elements that may be linked with this one to form an aggregate.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

partof: specifies the identifier of a **<join>** element in the current document representing an aggregate of which this element is a component part.
Data type: IDREF
Value: a valid SGML identifier attached to a **<join>** element.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

prev (i.e. previous): specifies the identifier of an element which may be linked with the current element to form an aggregate, and which immediately precedes it in the aggregate.
Data type: IDREF
Value: a valid SGML identifier for an element in the current document.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

next (i.e. next): specifies the identifier of an element which may be linked with the current element to form an aggregate, and which immediately follows it in the aggregate.
Data type: IDREF
Value: a valid SGML identifier for an element in the current document.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

Part: additional tag set for analysis and interpretation.
Member of classes:
DTD file:
Discussed in 16, “Segmentation and Alignment,” [in separate fascicle].

<respons> (i.e. responsibility)

Description: identifies the individual(s) responsible for some aspect of the markup of some particular element(s).
Attributes:

target: gives the SGML identiifer(s) of the element(s) for which some aspect of the responsibility is being assigned.
Data type: IDREFS
Value: one or more SGML identifiers.
Default value: #REQUIRED
Remarks:
This attribute is required.

locus: indicates the specific aspect of the markup for which responsibility is being assigned.
Data type: CDATA
Suggested values include:
#gi : responsibility for the claim that the element is of the type indicated by the markup
#location : responsibility for the claim that the element begins and ends where indicated
#startloc : responsibility for the claim that the element begins where indicated
#endloc : responsibility for the claim that the element ends where indicated
[attribute name] : responsibility for the claim that the attribute named has the value given in the markup
#transcribedContent : responsibility for the transcription of the element content
#suppliedContent : responsibility for the contents supplied by the encoder (corrections, expansions of abbreviations, etc.)
Default value: #REQUIRED

Example:

Remarks:
This attribute is required.

resp: identifies the individual or agency responsible for the indicated aspect of the electronic text.
Data type: CDATA
Value: any string of characters, typically the initials of an individual, the acronym of an agency, the name of a computer program, etc.
Default value: #REQUIRED
Remarks:
This attribute is required.

desc (i.e. description): gives a brief prose note supplying any additional information which should be recorded
Data type: CDATA
Value: any string of characters, typically a phrase or sentence in a natural language.
Default value: #IMPLIED
Remarks:
This attribute is optional.

Example:

Remarks: The **<resp>** element is designed for cases in which fine-grained information about specific aspects of the markup of a text is desirable for whatever reason. Global responsibility for certain aspects of markup is usually more simply indicated in the TEI header, using the **<resp>** element within the title statement, edition statement, or change log.
Part: additional tag set for.
Member of classes: global
DTD file:
Data description: Empty.
May occur within:
May contain: EMPTY
Declarations:

```
<!ELEMENT respons       - O  EMPTY                          >
<!ATTLIST respons            %a.global;
          target             IDREFS            #REQUIRED
          locus              CDATA             #REQUIRED
          resp               CDATA             #REQUIRED
          desc               CDATA             #IMPLIED       >
```

Discussed in 19.2, “Attribution of Responsibility,” on page 10.
This tag is optional.

Class Name: **spoken**

Description: defines additional attributes which are enabled for all elements when the base tag set for transcription of spoken texts is selected.

Attributes: [None: global and inherited attributes only.]

Part: additional tag set for spoken materials.

Member of classes:

DTD file:

Discussed in 11, “Base Tag Set for Transcriptions of Spoken Texts,” [in separate fascicle].

Class Name: **terminology** (i.e. global attributes for terminological data.)

Description: defines attributes for all elements in documents which use the base tag set for terminological data.

Attributes:

group: indicates the group (term and related elements) to which this element should be associated by specifying a string matching the *n* attribute value on an appropriate element.
Data type: CDATA
Value: any string matching the *n* attribute value on the **<term>** element to which the group is attached.
Default value: #IMPLIED
Example:

Remarks: The *group* attribute provides a specialized pointing mechanism for use within **<termEntry>** elements.
This attribute is optional.

grpPtr: indicates the group (term and related elements) to which this element should be associated by specifying its unique identifier, where this is available.
Data type: IDREF
Value: the value specified must match a value supplied as the value for an *id* attribute on some **<term>** element in the current SGML document.
Default value: #IMPLIED
Example:

Remarks: The *group* attribute provides a specialized pointing mechanism for use within **<termEntry>** elements.
This attribute is optional.

depend: indicates the parent element to which this element should be associated by specifying a string matching the *n* attribute value on an appropriate element.

Data type: CDATA
Value: any string matching the *n* attribute value on the element to which the dependent element is attached.
Default value: #IMPLIED
Example:

This attribute is optional.

depPtr: indicates the parent element to which this element should be associated by specifying its unique identifier, where this is available.
Data type: IDREF
Value: the value specified must match a value supplied as the value for an *id* attribute on some **<term>** element in the current SGML document.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

Remarks: The attributes shared by this element class are used for linking elements, possibly not adjacent in the record, which are related (e.g. a grammatical annotation and the term it describes). If no attribute is specified, the element is assumed to relate to the most recently specified **<term>** or **<otherForm>** element.
Part: base tag set for terminological data.
Member of classes:
DTD file:
Discussed in 14.2, “Tags for Terminological Data,” [in separate fascicle].
Discussed in 14.3.3, “Flat Term Entries Using Group and Depend Attributes,” [in separate fascicle].
Discussed in 14.4, “Overall Structure of Terminological Documents,” [in separate fascicle].

Usability Report and Comment

If you have comments of any kind on the current draft of the TEI Guidelines, please complete this form and return it to one of the following addresses:

C. M. Sperberg-McQueen, Computer Center (M/C 135), University of Illinois at Chicago, 1940 W. Taylor Room 124, Chicago IL 60612-7352, USA. Bitnet: U35395@UICVM, Internet: U35395@uicvm.uic.edu

Lou Burnard, Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN, UK. Fax: +44 (865) 273275, Janet: LOU@UK.AC.OXFORD.VAX Internet: LOU@vax.oxford.ac.uk

Further copies of this form may be downloaded from LISTSERV @ UICVM (or listserv @ uicvm.uic.edu). Send a mail message containing the line 'GET TEIURC MEMO TEI-L'

Your name:
Your postal address:

Your e-mail address (if any):
Your occupation:
Your academic background:
Your immediate reactions to the Guidelines (please tick or cross one only):

Relevance to your present concerns or interests: high medium low
Importance to your future research plans: high medium low
Comprehensibility/usability of current draft: high medium low

Detailed technical comment: [Please feel free to comment on any part of the Guidelines or in general. Comments relating to specific parts of the text are most helpful if they include a section or chapter number]

Thank you!

Temporary Index