

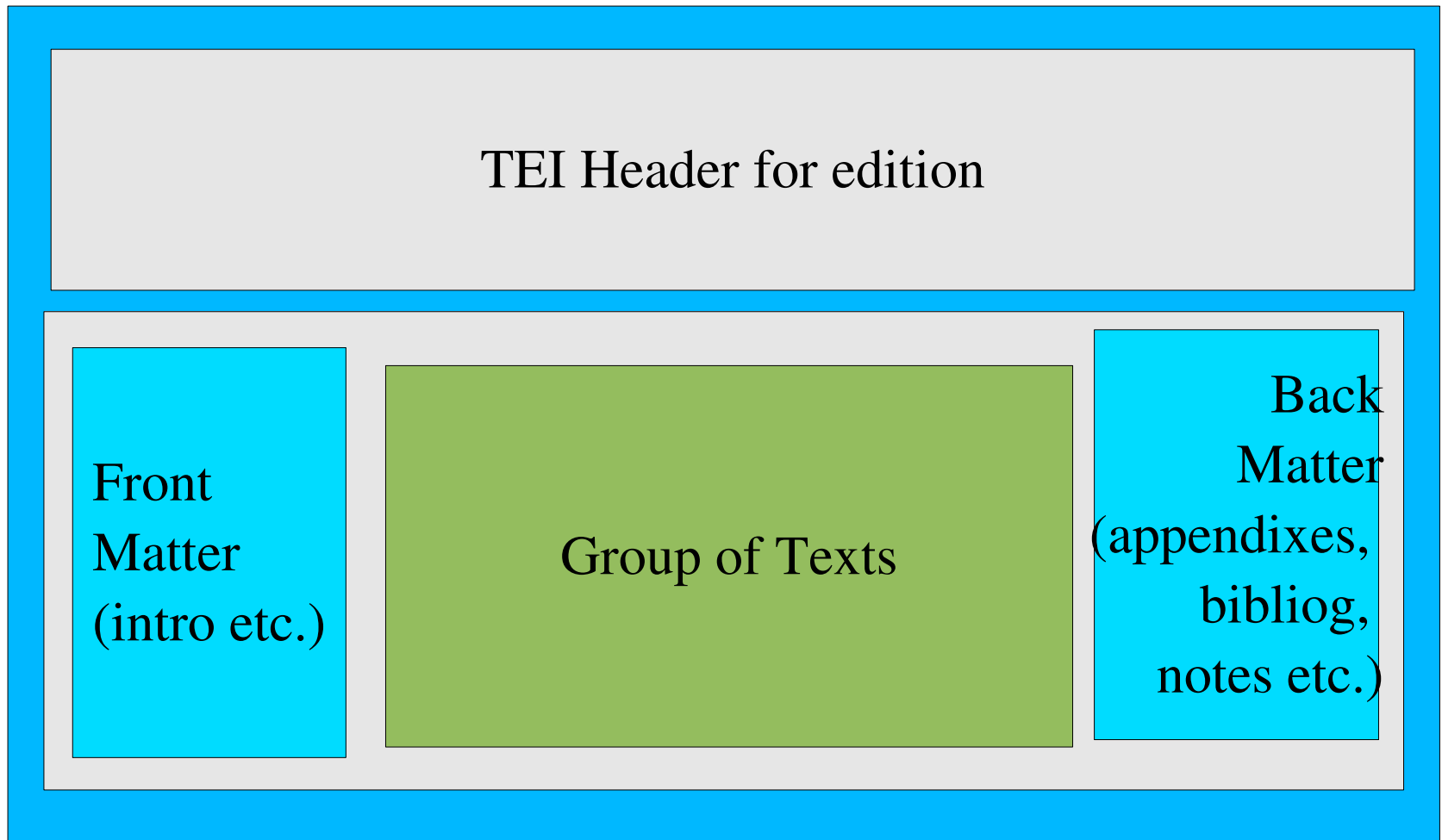
an example project: EETS

- ★ Background
 - ★ Early English Text Society
 - ★ Legacy conversion problem
 - ★ Pilot digitization project by Bella Willet
- ★ <http://www.tei-c.org/Projects/EETS/>
- ★ The *Ancrene Wisse*

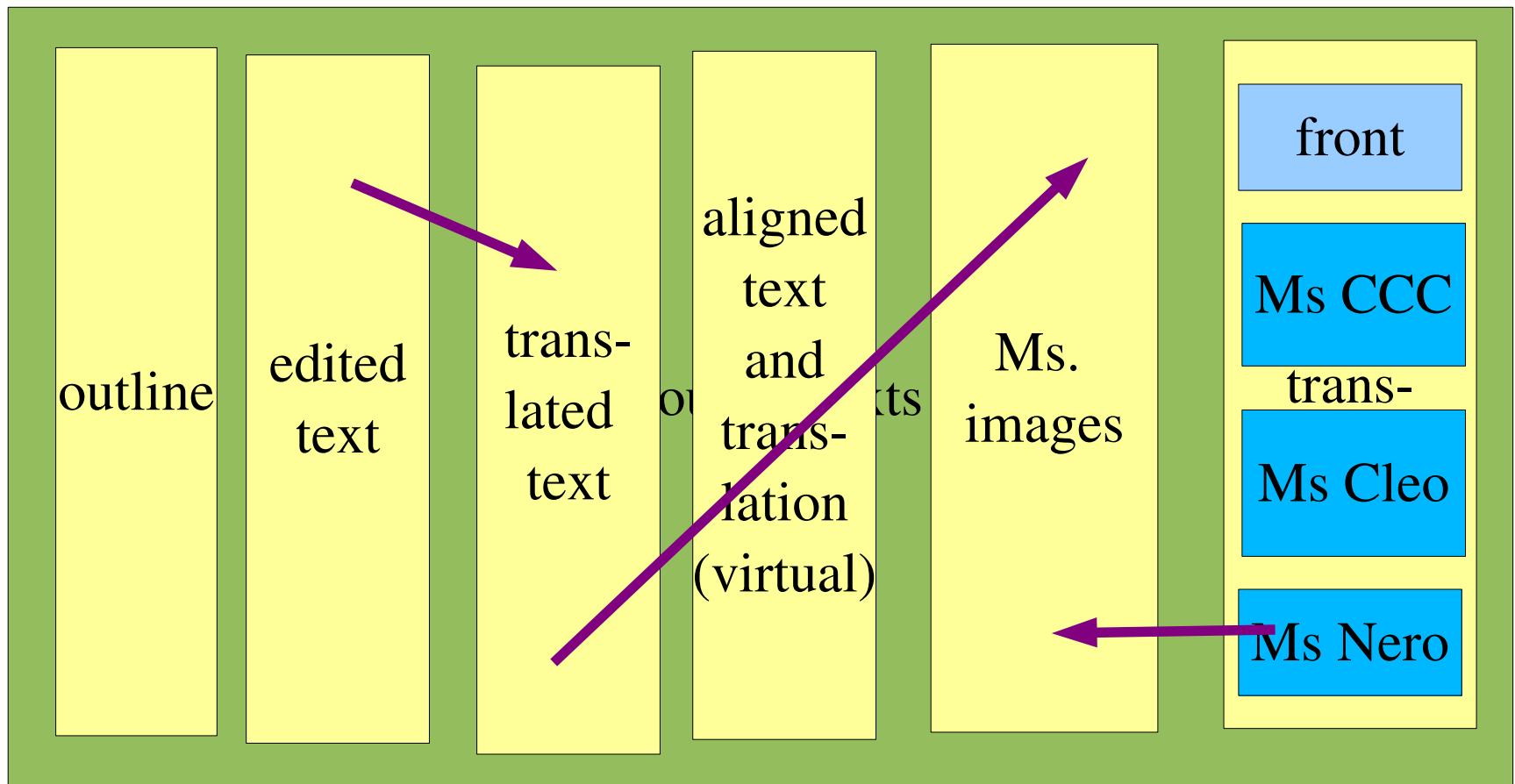
EETS components

- ★ Manuscript page images
- ★ Annotated transcriptions
- ★ Critical (synthetic) edition
- ★ Modern translation and summary
- ★ Notes, glossary, foreword, bibliography, etc.
- ★ Manuscript descriptions and metadata

EETS structure



EETS structure, cont.



<p><pp target="P1"/> <lab
<lat>Recti diligunt te</l
three kinds of <q>rightne
theological, the third co
disposition and external

<p id="P1"> <lat>R
sponsum. Est rectum
theologicum; et sunt
theologico sermo no
directionem, altera u
rectificationem.</lat></p>

<page id="A-f1r">
<pp target="P1"/> <lb/>
feaderes <expan>ant</ex
<lb/>IN canticis. spon=
gastes nome</title>
<lb/>sa ad sponsum. <ti
<lb/>Est rectum gramati
<title rend="rubric">.w
<lb/>geometricum. Rectu
<lb/>Et sunt difference
theolo

<div rend="notoc" n="A">

<head>

<ref target="AW-MS-A-desc">

Cambridge, Corpus Christi College, MS 402

</ref></head>

<p>

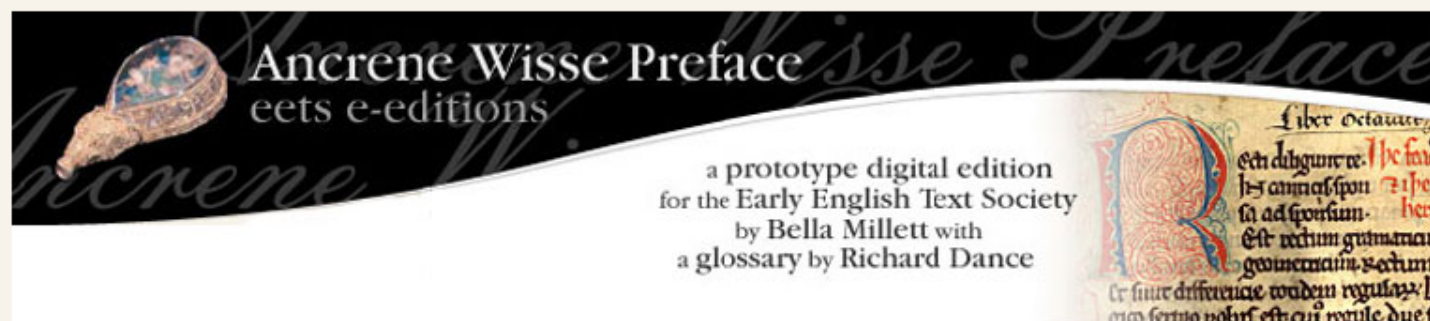
<figure n="f1r" corresp="A-f1r"/>

<figure n="f1v" corresp="A-f1v"/>



EETS implementation

- ★ Uses TEI modules for linking, transcription, figures
- ★ Adds new elements for e.g. latin, middle-english, and ornamental initial letters
- ★ XSLT stylesheet converts to HTML, with embedded links
- ★ CSS defines appearance



Search
Home
Editor's preface
Introduction
Outline summary
Edited text
Translated text
Text and translation
MS Reproductions
MS Transcriptions
Apparatus criticus
Textual Commentary
Glossary
Booklist

About the project

This edition was developed by [Bella Millett](#) for the [Early English Text Society](#) with technical assistance from Lou Burnard and Sebastian Rahtz of the [Text Encoding Initiative](#), and design input from Scott Agass of [Next New Media](#). It was supported by a grant from the [AHRB](#) Resource Enhancement Scheme.

Editor's preface



The Editor's Preface gives an account of the background and aims of the project, an introduction to the use of the website, and a review of the problems involved in e-editing.

[read more](#)

Introduction



The Introduction deals with the edition itself: background information on *Ancrene Wisse* and its Preface, descriptions of the MSS, and an account of the principles and conventions of editing.

[read more](#)



Cambridge,
Corpus Christi College,
MS 402, f. 1r

Home | Editor's preface | Introduction | Outline summary | Edited text | Translated text | **Text and translation** | MS Reproductions | MS Transcriptions | Apparatus criticus | Textual Commentary | Glossary | Booklist

Text and translation

Text

Translation

2 I þe Feaderes ant i þe Sunes ant i þe Hali Gastes
nome, her biginneð *Ancrone Wisse*.

In the name of the Father, and of the Son, and of
the Holy Ghost, the *Guide for Anchoresses* begins
here.

[1] *Recti diligunt te. In Canticis: sponsa ad
sponsum. Est rectum gramaticum, rectum
geometricum, rectum theologicum; et sunt
differencie totidem regularum. De recto theologico
sermo nobis est, cuius regule due sunt: vna circa
cordis directionem, altera uersatur circa exteriorum
rectificationem.*

'The upright love you.' In the Song of Songs: the
bride to the bridegroom. There is what is right in
grammar, what is right in geometry, and what is right
in theology; and each of these is a different kind of
rightness. We are talking about rightness in a
theological sense, which has two rules: one is
concerned with the direction of the heart, the other
with the rectification of external things.

[2] *Recti diligunt te.* 4 'Loverd' seið Godes spuse to *The upright love you. 'Lord' says the bride of God*

Edited text

Search

Home

Editor's preface
IntroductionOutline summary
Edited text

– Notes

Translated text
Text and translation
MS Reproductions
MS TranscriptionsApparatus criticus
Textual Commentary
Glossary
Booklist2 | [pe Feaderes ant i pe Sunes ant i pe Hali Gastes nome, her biginneð Ancrene Wisse.](#)

[1] *Recti diligunt te. In Canticis: sponsa ad sponsum. Est rectum gramaticum, rectum geometricum, rectum theologicum; et sunt difference totidem regularum. De recto theologico sermo nobis est, cuius regule due sunt: vna circa cordis directionem, altera uersatur circa exteriorum rectificationem.*

[Translation](#) | [Text + Translation](#) | [Apparatus Criticus](#) | [Textual Commentary](#)

[2] *Recti diligunt te.* 4 ‘Lauerd’, seið Godes spuse to hire deore[wu]rðe spus, ‘pe rihte luuieð pe.’ Peo beoð rihte pe 5 l[i]uieð efter riwle. Ant 3e, mine leoue sustren, habbe[ð] moni dei icrauet on me [e]fter ri[w]le. 6 Monie cunne riwle beoð; ah twa beoð bimong alle þet Ich chulle spoken of þurh ower bone, 7 wið Godes grace.

[Translation](#) | [Text + Translation](#) | [Apparatus Criticus](#) | [Textual Commentary](#)

[3] 8 Pe an riwleð pe heorte¹, ant makeð efne ant smeðe wiðute cnost ant dolc of woh inwit ant of 9 wreiðende þe segge, ‘Her þu sunegest’, oper ‘Pis nis nawt ibet 3et as wel as hit ahte.’ Peos riwle 10 is eauer inwið ant rihteð pe heorte. *Et hec est caritas quam describit Apostolus, de corde puro et consciencia bona et fide non ficta.* Peos riwle is chearite of schir heorte ant cleane inwit ant 11 treowe bileaue. *Pretende, inquit Psalmista, misericordiam tuam scientibus te per fidem non fictam, et iusticiam tuam—id est, uite rectitudinem—hiis qui recto sunt corde—qui scilicet omnes uoluntate suas dirigunt ad regulam diuine uoluntatis. Isti dicuntur boni anto[no]masice: Psalmista: Benefac, Domine, bonis et rectis corde. Isti dicitur ut glorientur—testimonio uidelicet bone conscientie: Gloriamini, omnes recti corde, quos scilicet rectificauit regula illa suprema rectificans omnia, de qua Augustinus: Nichil petendum preter regulam magisterii, et Apostolus: Omnes in eadem regula permaneamus.*

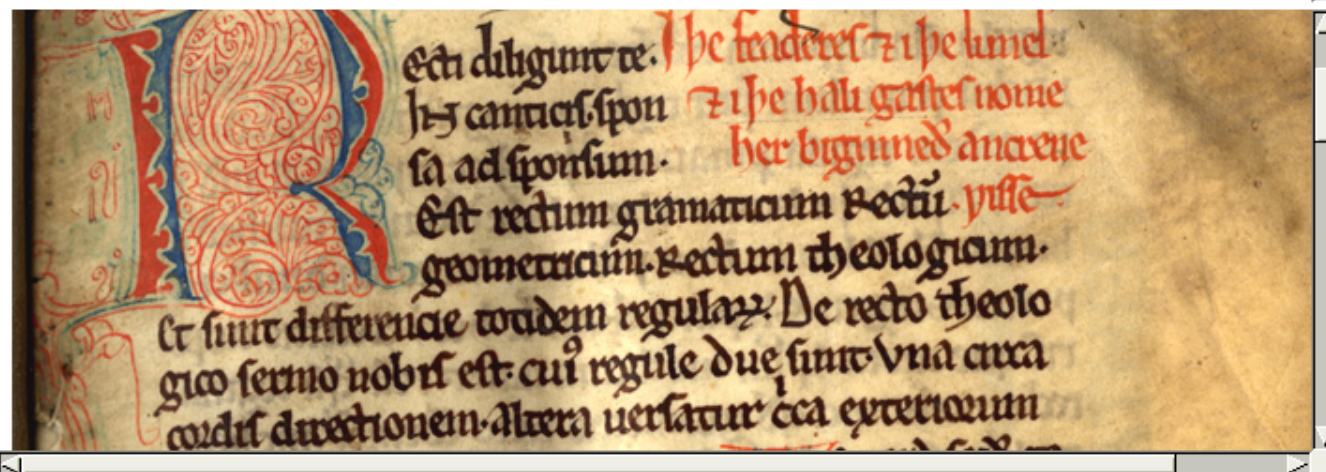
[Translation](#) | [Text + Translation](#) | [Apparatus Criticus](#) | [Textual Commentary](#)

Cambridge, Corpus Christi College, MS 402 | MS C (Cleopatra C. vi) | MS N (Nero A. xiv)

f1r | f1v | f2r | f2v | f3r | f3v | f4r | f4v

[Transcribed text]

- [1]
 1 Recti diligunt te. I þe feaderes *ant* i þe sunes
 2 IN canticis. spon= *ant* i þe hali gastes nome
 3 sa ad sponsum. her biginneð ancrene
 4 Est rectum gramaticum Rectum .wisse.
 5 geometricum. Rectum theologicum.
 6 Et sunt difference totidem regularum. De recto theolo=
 7 gico sermo nobis est. cuius regule due sunt. Vna circa
 8 cordis directionem. Altera uersatur circa exteriorum
 9 rectificationem. [2] Recti diligunt te. Lauerd seið go=
 10 des spuse to hire deorewurðe¹ spus. þe rihte luueð þe.
 11 þeo beoð rihte þe luueð efter riwle. Ant 3e mine leo=
 12 ue sustren habbeð moni dei icrauet on me efter riwle. 2
 13 Monie cunne riwlen beoð. ah twa beoð bimong alle.
 14 þe¹ ich chulle spoken of þurh ower bone wið godes
 15 grace. [3] þe an riwleð þe heorte *ant* makeð efne *ant* smeðe wið=



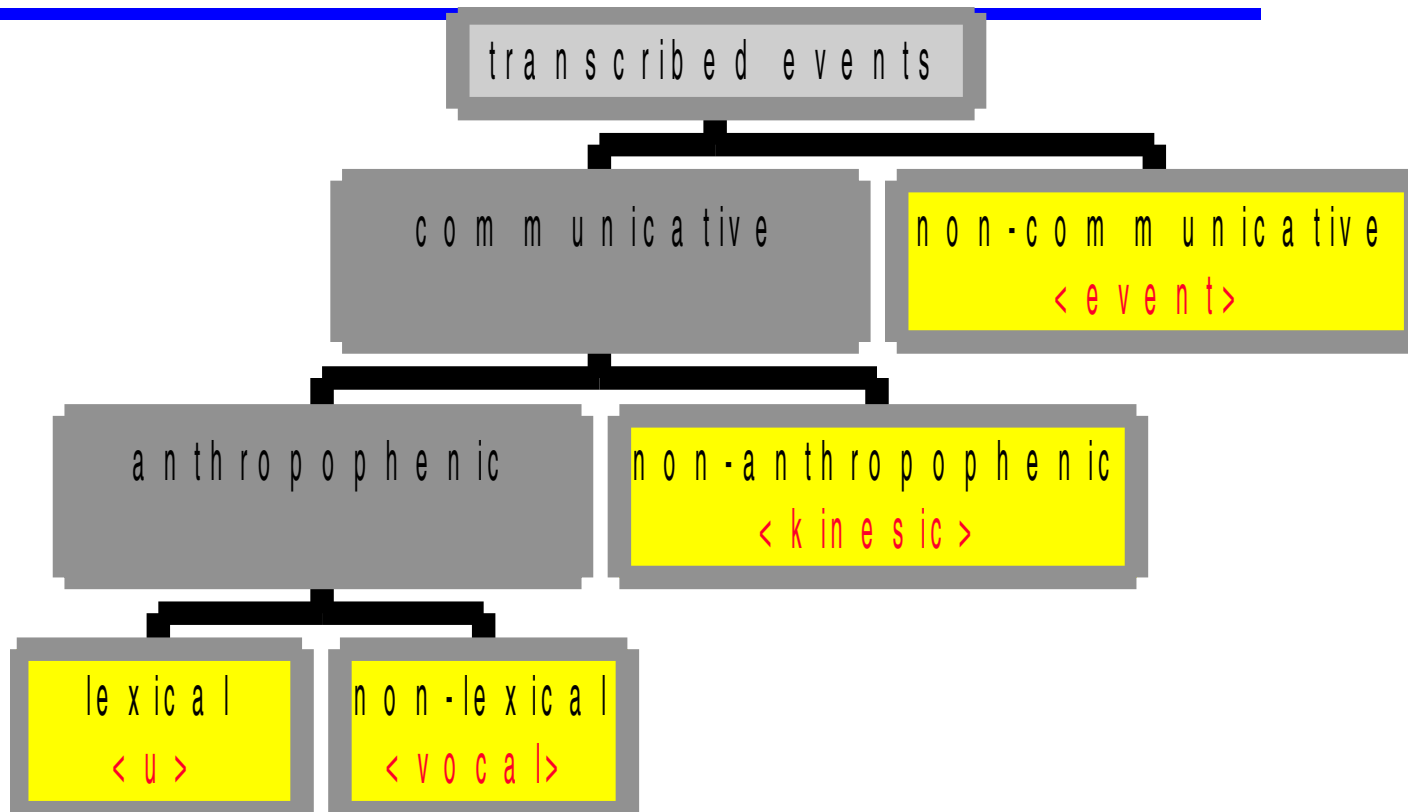
Transcribing speech

- ★ normalization issues
- ★ ease of reading vs accuracy
- ★ interpretation vs prosody
- ★ analogous to problems of handling digitized images

The Spoken base tagset

- ★ components : **<u>** **<event>** **<kinesic>** **<vocal>**
<pause> **<shift>**
- ★ contextual information in header **<settingDesc>**
<particDesc>
- ★ facilities for synchronization and timing

Features of speech



Utterances

- ★ Basic unit of discourse, corresponding to speaker turns
- ★ Optionally grouped into higher-level divisions (<div>s), e.g. to mark discourse function
- ★ Linked by **who** attribute to <person> description in header

Vocals and events

- ★ Empty elements are used to mark paralinguistic phenomena

```
<u who="Jan">This is just delicious</u>
<event desc='telephone rings'>
<u who="Kim">I'll get it</u>
<u who="Tom">I used to <vocal desc="cough"/>
smoke a lot</u>
<u who="Bob"><vocal desc="sniff"/>He thinks
he's tough</u>
<vocal who="Ann" desc="snorts"/>
```

Voice quality and prosody

- ★ The **<shift>** element is used to mark changes in voice quality

```
<u who="LB">  
<shift feature="loud" new="f"/>Elizabeth</u>  
<u who="EB">Yes</u>  
<u who="LB"><shift/>Come and try this  
<pause/>  
<shift feature="loud" new="ff"/>come on</u>
```

- ★ Other prosodic features may be marked using specific kinds of **<seg>** or entity refs

Another example

```
<u who="MAR">you never <pause/> take this cat for  
show and tell <pause dur='5' /> meow meow</u>  
<u who="ROS">yeah well I dont want to</u>  
<event><desc>toy cat has bell in tail which  
continues to make a tinkling sound</desc></event>  
<vocal who="MAR"><desc>miaows</desc></vocal>  
<u who="ROS">because it is so old</u>  
<u who="MAR">how <reg>about</reg> your cat <pause/>  
yours is new  
<kinesic><desc>shows Father the  
cat</desc></kinesic></u>  
<u who="FAT" trans="pause">that<pause/> darling</u>  
<u who="MAR"><s>no mine isnt old</s>  
  <s>mine is just um a little dirty</s></u>
```

Participant Description

```
<person xml:id="P1" sex="2" age='mid'>
<birth date='1950-01-12'>
  <date>12 January 1950</date>
  <name type="place">Shropshire, UK</name>
</birth>
<firstLang>English</firstLang>
<langKnown>French</langKnown>
<residence>Long term resident of
Hull</residence>
<education>University postgraduate</education>
<occupation>Unknown</occupation>
<socecstatus source="PEP" code="B2"/>
</person>
```

Setting Description

```
<settingDesc>
<setting who="#P1 #P2"><name type="city">Bedford</>
<name type="region">UK: South East</name>
<date value="1989">early spring, 1989</>
<locale>rug of a suburban
home</locale><activity>playing</activity></setting>
<setting who="#P3"><name type="city">Bedford</name><name
type="region">UK: South East</name><date
value="1989">early spring, 1989</date><locale>at the
sink</locale> <activity>washing-up</activity></setting>
<setting who="#P4"><name type="place">London, UK</name>
<time>unknown</time><locale>broadcasting studio</locale>
<activity>radio performance</activity>
</setting></settingDesc>
```

Timing

- ★ Pausing
 - ★ use **<pause>** element
- ★ Duration
 - ★ use **dur** attribute
- ★ Overlap
 - ★ use **trans** attribute

Overlap

Have you heard the
the election results?

its a disaster
its a miracle

```
<u xml:id="A1" who="A">Have you heard the</u>  
<u xml:id="B1" who="B" trans="latching">the election  
results? </u>  
<u xml:id="A2" who="A" trans="pause">its a  
disaster</u>  
<u xml:id="B2" who="B" trans="overlap">its a miracle  
</u>
```

The Dictionary base tagset

- ★ primarily for printed dictionaries, rather than lexica or dictionary production systems
- **<entry>**, **<entryFree>**, and **<superEntry>**
- **<sense>** and **<hom>**
- ★ logical structure vs. typographic fidelity

Constituents of a Dictionary Entry

- ★ the form group
- ★ the grammatical-information group
- ★ the definition or translation
- ★ etymology
- ★ examples
- ★ usage information
- ★ cross-references to other entries
- ★ notes and related entries

Dictionary components (1)

- **<form>** grouping element for one or more of **<orth>** **<pron>** **<hyph>** **<syll>** **<stress>** etc.
- **<gramGrp>** groups specialised grammatical tags **<gen>**, **<number>** etc
- **<def>** for definition text, **<trans>** for translation
- **<etym>** for etymology

Dictionary components (2)

- ★ examples **<eg>**
- ★ usage note **<usg>**
- ★ label **<lbl>**
- ★ related entries **<re>** and specialized pointers **<oRef>**, **<pRef>** etc

Simple example

```
<entry><form><orth>OATS,</orth>  
<gram>n. s.</gram>  
<etym>[aten, Sax.]</etym>  
<def>A grain, which in England is  
generally given to horses; but in  
Scotland supports the people.</def>  
</form></entry>
```

Linking, segmentation, alignment

- ★ Provides generic segmentation elements
- ★ Provides extensive set of attributes for linkage, correspondence, synchronization, aggregation, alternation, etc.
- ★ Documents generic pointing mechanism

Generic segmentation elements

- **<seg>** for arbitrary (nesting) segmentation
- **<s>** for end-to-end segmentation
- ★ use **type** attribute to subcategorise
- **<anchor>** for points
- ★ Segmentation is the key to successful linking and analysis

Clustering

(Difficulty (is being expressed)
with ((the method) (to be used)))

```
<s>Difficulty <seg>is being  
expressed</seg> with  
<seg><seg>the method</seg>  
<seg>to be  
used</seg></seg></s>
```

discontinuous segments

“You put it,” Quill reminded him, “in the safe.”

```
<s xml:id="s1">"You put it,"</s>  
<s xml:id="s2">Quill reminded  
him,</s>  
<s xml:id="s3">"in the safe."</s>
```

- ★ fundamental problem
- ★ first segment, then link, using stand-off

discontinuous segments

“You put it,” Quill reminded him, “in the safe.”

```
<s xml:id="s1" next="#s3">"You put it,"</s>  
<s xml:id="s2">Quill reminded him,</s>  
<s xml:id="s3" prev="#s1">"in the safe."</s>
```

- ★ can also use **PART** attribute to indicate that segments are incomplete

discontinuous segments

“You put it,” Quill reminded him, “in the safe.”

```
<s xml:id="s1">"You put it,"</s>  
<s xml:id="s2">Quill reminded him,</s>  
<s xml:id="s3">"in the safe."</s>
```

```
<join targets="#s1 #s3" result="s"/>
```

Anaphoric reference

```
<title xml:id="shirl">Shirley</title>,  
which made its Friday night debut only a  
month ago, was not listed on <name  
xml:id="nbc">NBC</name>'s new schedule,  
although <seg xml:id="nwk"  
corresp="#nbc">the network</seg> says  
<seg xml:id="show" corresp="#shirl">the  
show</seg> still is being considered.
```

```
<correspGrp type="anaphor">  
  <link targets="#shirl #show"/>  
  <link targets="#nwk #nbc"/>  
</correspGrp>
```

Translation pairs

```
<s xml:id="s1" corresp="#s2"  
xml:lang="EN">
```

For a long time I used to go to bed
early</s>

```
<s xml:id="s2" corresp="#s1"  
xml:lang="FR">
```

Longtemps je me couchais de bonne
heure</s>

and/or....

```
<correspGrp type="trans">  
  <link targets="#s1 #s2"/>  
</correspGrp>
```

Synchronization

★ of whole elements

```
<u xml:id="A2" who="A" synch="#B2">  
its a disaster</u>  
<u xml:id="B2" who="B">its a  
miracle</u>
```

★ of points in time

```
<u xml:id="A1" who="A">Have you heard  
<anchor xml:id="A0"/>the</u>  
<u xml:id="B1" who="B" start="#A0">  
<anchor xml:id="B0"/>the election  
results? yes</u>
```

Analytic mechanisms

- ★ Specific kinds of segment for linguistic analyses
- ★ Specialized interpretive pointers (**** and **<spanGrp>**)
- ★ The **ana** attribute and its possible targets
 - **<interp>** and **<interpGrp>**
 - ★ feature systems **<fs>** and **<fsd>**

Arbitrary characterizations

- ★ The **** points into a stretch of a text and characterizes it in some way
- ★ Target may be anything you can reach by an xpath

```
<spanGrp resp="#LB" type="thematic" >  
  <span value="ships" from="#P1" to="#P2"/>  
  <span value="shoes" from="#P4" to="#P8"/>  
  <span value="sealing wax"  
from="http://www.somewhere.com/waxinit.xml#P45"/>  
</spanGrp>
```

More detailed analysis

```
<w ana="#VVD">annotated</w>
```

- ★ the **ana** attribute is of type IDREFS
- ★ what does **VVD** identify?
 - ★ a prose description
 - ★ an **<interp>** element
 - ★ a feature structure

using interp...

```
<w ana="#VVD">annotated</w>  
<w ana="#NN2">corpora</w>
```

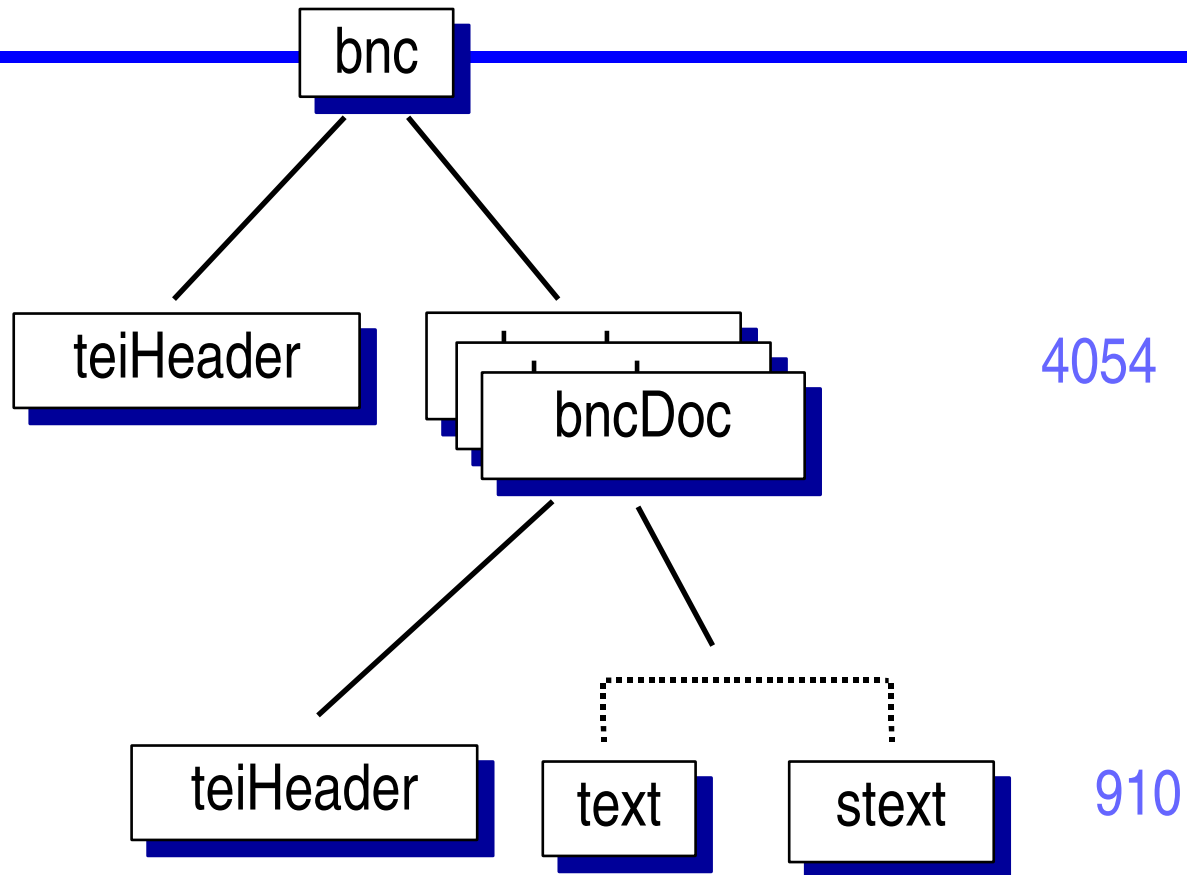
```
<interp xml:id="VVD"  
type="lexicalClass"  
value="verbPastTense"/>  
<interp xml:id="NN2"  
type="lexicalClass"  
value="nounPlural"/>
```

hierarchic grouping of interps

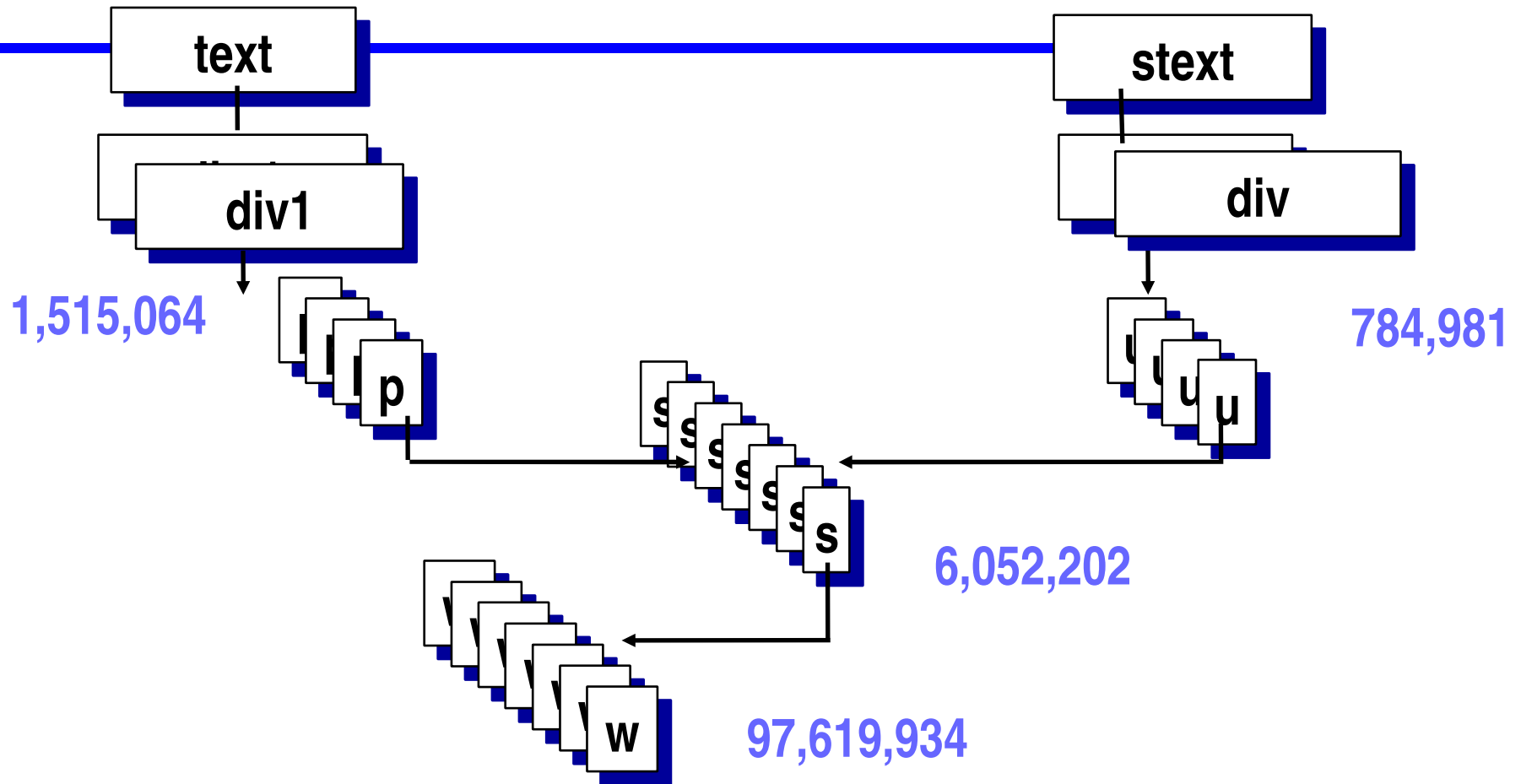
- ★ nouns can be common or proper
- ★ nouns can be singular or plural

```
<interpGrp value="nomimal">  
  <interpGrp value="common">  
    <interp value="singular"/>  
    <interp value="plural"/>  
  </interpGrp>  
</interpGrp>
```

BNC structure



BNC structure



Word class annotation

- ★ CLAWS (Leech, Garside *et al*) approach
- ★ What counts as a word?

This isn't *prima facie* obvious, in spite of spelling conventions.

- ★ What counts as a word class?

NN1 NN2 NP1 NP2 T00. . .

Representation of annotation

```
<s n="00011">  
  <w ana="AT0">The</w>  
  <w ana="NP0">Queen</w>  
  <w ana="POS">'s</w>  
  <w ana="AJ0">real</w>  
  <w ana="NN1">annus horribilis</w>  
  <w ana="VVD">began</w>  
  <w ana="PRP">on</w>  
  <w ana="NP0-NN1">Sunday  
  <c ana="PUN">.</c>  
</s>
```