Future Learn

Corpus Linguistics

1. Introduction to Corpora

Some untagged text

"Arrest warrant out for Clowes' partner years before collapse."
By Daniel John

A WARRANT for the arrest of the former partner of Mr Peter Clowes was issued seven years before his Barlow Clowes investment empire collapsed, according to evidence submitted to the Parliamentary Ombudsman.

Add tags for headers and paragraphs

<head type=MAIN>
"Arrest warrant out for Clowes' partner years before collapse."
</head>
<head type=BYLINE>
By Daniel John
</head>
<p>
A WARRANT for the arrest of the former partner of Mr Peter Clowes was issued seven years before his Barlow Clowes investment empire collapsed, according to evidence submitted to the Parliamentary Ombudsman.
</p>

Add sentence tags

\<head type=MAIN\>

\<s n=001\>"Arrest warrant out for Clowes' partner years before collapse."

\</head\>

\<head type=BYLINE\>

\<s n=002\>By Daniel John

\</head\>

\<p\>

\<s n=003\>A WARRANT for the arrest of the former partner of Mr Peter Clowes was issued seven years before his Barlow Clowes investment empire collapsed, according to evidence submitted to the Parliamentary Ombudsman.

\</p\>

Change quotes to SGML

<head type=MAIN>

<s n=001>&bquo;Arrest warrant out for Clowes' partner years before collapse&equo;

</head>

<head type=BYLINE>

<s n=002>By Daniel John

</head>

<p>

<s n=003>A WARRANT for the arrest of the former partner of Mr Peter Clowes was issued seven years before his Barlow Clowes investment empire collapsed, according to evidence submitted to the Parliamentary Ombudsman.

</p>

Add tags for punctuation

<head type=MAIN>

<s n=001><c PUQ>&bquo;Arrest warrant out for Clowes<c PUN>' partner years before collapse <c PUQ>&equo;

</head>

<head type=BYLINE>

<s n=002>By Daniel John

</head>

<p>

<s n=003>A WARRANT for the arrest of the former partner of Mr Peter Clowes was issued seven years before his Barlow Clowes investment empire collapsed, according to evidence submitted to the Parliamentary Ombudsman <c PUN>.

</p>

Add tags for word units

<head type=MAIN>
<s n=001><c PUQ>&bquo<w>Arrest <w>warrant <w>out <w>for <w>Clowes<c PUN>'
  <w>partner <w>years <w>before <w>collapse <c PUQ>&equo'
<c PUN>.
</head>
<head type=BYLINE>
<s n=002><w>By <w>Daniel <w>John
</head>
<p>
<s n=003><w>A <w>WARRANT <w>for <w>the <w>arrest <w>of <w>the <w>former
  <w>partner <w>of <w>Mr <w>Peter <w>Clowes <w>was <w>issued <w>seven
  <w>years <w>before <w>his <w>Barlow <w>Clowes <w>investment <w>empire
  <w>collapsed<c PUN>, <w>according to <w>evidence <w>submitted <w>to
  <w>the <w>Parliamentary <w>Ombudsman<c PUN>.
</p>

Add grammatical codes to word units

<head type=MAIN>
<s n=001><c PUQ>&bquo<w NN1>Arrest <w NN1>warrant <w AVP>out <w PRP>for <w
    NP0>Clowes<c PUN>' <w NN1>partner <w NN2>years <w PRP>before <w NN1>collapse <c
    PUQ>&equo'
<c PUN>.
</head>
<head type=BYLINE>
<s n=002><w PRP>By <w NP0>Daniel <w NP0>John
</head>
<p>
<s n=003><w AT1>A <w=NN1>WARRANT <w PRP>for <w AT0>the <w NN1>arrest <w PRF>of <w
    AT0>the <w DT0>former <w NN1>partner <w PRF>of <w NP0>Mr <w NP0>Peter <w
    NP0>Clowes <w VBD>was <w VVN>issued <w CRD>seven <w NN2>years <w CJS>before <w
    DPS>his <w NN1-NP0>Barlow <w NP0>Clowes <w NN1>investment <w NN1>empire <w
    VVD>collapsed<c PUN>, <w PRP>according to <w NN1>evidence <w VVN>submitted <w
    PRP>to <w AT0>the <w AJ0>Parliamentary <w NN1>Ombudsman<c PUN>.
</p>