

Link to the Balisage Proceedings Home Page at <https://www.balisage.net/Proceedings>

Link to this paper (entry page with links) at <https://www.balisage.net/Proceedings/vol30/html/Pemberton01/BalisageVol30-Pemberton01.html>



Balisage Paper: The Book of Doublends Jined: Parsing *Finnegans Wake* with ixml

Steven Pemberton

Researcher

CWI, Amsterdam

<steven.pemberton@cwi.nl>

Balisage: The Markup Conference 2025

August 4 - 8, 2025

© Steven Pemberton, 2025

How to cite this paper

Pemberton, Steven. "The Book of Doublends Jined: Parsing *Finnegans Wake* with ixml." Presented at Balisage: The Markup Conference 2025, Washington, DC, August 4 - 8, 2025. In *Proceedings of Balisage: The Markup Conference 2025*. Balisage Series on Markup Technologies, vol. 30 (2025). <https://doi.org/10.4242/BalisageVol30-Pemberton01>.

Abstract

Finnegans Wake by James Joyce is probably the hardest book to read in the English language. A principle hurdle is the length and convolutedness of the sentences. This paper reports work-in-progress of an attempt to handle the complexity of *Finnegans Wake* by parsing the sentences (at a structural, not a semantic level), to reveal their top-level structure. It takes the reader step-by-step through the construction of an ixml grammar for dealing with one chapter of the book.

► Table of Contents

Introduction

Finnegans Wake [fw] is probably the hardest book to read in the English language, if indeed you can say it is in English at all. There are three things that make it hard. Firstly, it's Joyce: it was not his aim to make books easy to read. The best music is that that you have to listen to several times before you get it; so it is with Joyce. He doesn't lead you gradually in, but throws you fully clothed into the deep end, to sink or swim. And he uses punctuation sparingly: for instance, he doesn't use quotation marks to tell you when someone is speaking, nor does he always tell you who is speaking.

Secondly, with *Finnegans Wake*, there's the vocabulary, where Joyce uses a lot of his own made-up portmanteau words and puns. When he refers to *Finnegans Wake* as "The Book of Doublends Jined" he's punning on "Dublin's Giant" and "Double Ends Joined", since Finn is a giant said to be buried under Dublin, and the two ends of the book join up to make a complete sentence, and thus a circular book. Or with the description of the wake for the dead giant: "And the all gianed in with the shoutmost shoviality." The underlying meaning is "They all joined in with the utmost joviality", with the added elements of giant,

shouting, and shoving, and probably the slurred speech of people who have had a little too much to drink. It is adroit use of language to express so much with so few words.

And thirdly, and harder yet, is the length and structure of the sentences. In the book, on page 122, when *Finnegans Wake* is being self-referentially described, there is a reference to the TUNC page of the *Book of Kells* [kells], an Irish mediaeval illustrated manuscript. This page, that begins with the word TUNC, contains but a single sentence (in Latin) that reads "Then they crucified Christ with two thieves." But the sentence is formatted in the form of a large X, and there are numerous adornments all round the page. Joyce's sentences are a bit like that: there is a simple essence, but he has added enormous amounts of symbolic adornments to them, making them hard to decypher at a first reading. And they often take up a whole page.

As an example of this, take this particularly long sentence from page 38 of the original:

Our cad's bit of strife (knee Bareniece Maxwellton) with a quick ear for spittoons (as the aftertale hath it) glaned up as usual with dumbestic husbandry (no persicks and armelians for thee, Pomeranzia!) but, slipping the clav in her claw, broke of the matter among a hundred and eleven others in her usual curtsey (how faint these first vhespers womanly are, a secret pispigliando, amad the lavurdy den of their manfolker!) the next night nudge one as was Hegesippus over a hup a ' chee, her eys dry and small and speech thicklish because he appeared a funny colour like he couldn't stood they old hens no longer, to her particular reverend, the director, whom she had been meaning in her mind primarily to speak with (hosch, intra! jist a timblespoon!) trusting, between cuppled lips and annie lawrie promises (mighshe never have Esnekerry pudden come Hunanov for her pecklapitschens!) that the gossip so delivered in his epistolear, buried teatoastally in their Irish stew would go no further than his jesuit's cloth, yet (in vinars venitas! volatiles valetotum!) it was this overspoiled priest Mr Browne, disguised as a vincentian, who, when seized of the facts, was overheard, in his secondary personality as a Nolan and underreared, poul soul, by accident — if, that is, the incident it was an accident for here the ruah of Ecclectiastes of Hippo outpuffs the writress of Havvah-ban-Annah — to pianissime a slightly varied version of Crookedribs confidentials, (what Mère Aloyse said but for Jesuphine's sake!) hands between hahands, in fealty sworn (my bravor best! my fraur!) and, to the strains of The Secret of Her Birth, hushly pierce the rubiend aurellum of one Philly Thurnston, a layteacher of rural science and orthophonethics of a nearstout figure and about the middle of his forties during a priestly flutter for safe and sane bets at the hippic runfields of breezy Baldoyle on a date (W. W. goes through the cald) easily capable of remembrance by all pickers-up of events national and Dublin details, the doubles of Perkin and Paullock, peer and prole, when the classic Encourage Hackney Plate was captured by two noses in a stablecloth finish, ek and nek, some and none, evelo nevelo, from the cream colt Bold Boy Cromwell after a clever getaway by Captain Chaplain Blount's roe hinny Saint Dalough, Drummer Coxon, nondepict third, at breakneck odds, thanks to you great little, bonny little, portey little, Winny Widger! you're all their nappies! who in his never-rip mud and purpular cap was surely leagues unlike any other phantomweight that ever toppitt our timber maggies.

The essence of this sentence is:

Our cad's bit of strife (knee Bareniece Maxwellton) with a quick ear for spittoons (as the aftertale hath it) glaned up as usual with dumbestic husbandry (no persicks and armelians for thee, Pomeranzia!) but, slipping the clav in her claw, **broke of the matter** among a hundred and eleven others in her usual curtsey (how faint these first vhespers womanly are, a secret pispigliando, amad the lavurdy den of their manfolker!) **the next night nudge one** as was Hegesippus over a hup a ' chee, her eys dry and small and speech thicklish because he appeared a funny colour like he couldn't stood they old hens no longer, **to her particular reverend**, the director, whom she had been meaning in her mind primarily to speak with (hosch, intra! jist a timblespoon!) **trusting**, between cuppled lips and annie lawrie promises (mighshe never have Esnekerry pudden come Hunanov for her pecklapitschens!) **that the**

gossip so delivered in his epistolear, buried teatoastally in their Irish stew **would go no further than his jesuit's cloth, yet** (in vinars venitas! volatiles valetotum!) **it was this overspoiled priest Mr Browne**, disguised as a vincentian, **who**, when seized of the facts, **was overheard**, in his secondary personality as a Nolan and underreared, poul soul, by accident — if, that is, the incident it was an accident for here the ruah of Ecclectiastes of Hippo outpuffs the writress of Havvah-ban-Annah — **to pianissime a slightly varied version** of Crookedribs confidentials, (what Mère Aloyse said but for Jesuphine's sake!) hands between hahands, in fealty sworn (my bravor best! my fraur!) **and**, to the strains of The Secret of Her Birth, **hushly pierce the rubiend aurellum of one Philly Thurnston, a layteacher of rural science** and orthophonethics of a nearstout figure and about the middle of his forties **during a priestly flutter** for safe and sane bets **at the hippic runfields** of breezy Baldoyle on a date (W. W. goes through the cald) easily capable of remembrance by all pickers-up of events national and Dublin details, the doubles of Perkin and Paullock, peer and prole, when the classic Encourage Hackney Plate was captured by two noses in a stablecloth finish, ek and nek, some and none, evelo nevelo, from the cream colt Bold Boy Cromwell after a clever getaway by Captain Chaplain Blount's roe hinny Saint Dalough, Drummer Coxon, nondepict third, at breakneck odds, thanks to you great little, bonny little, portey little, Winny Widger! you're all their nappies! who in his never-rip mud and purpular cap was surely leagues unlike any other phantomweight that ever toppitt our timber maggies.

and the distillation thus:

Our cad's bit of strife broke of the matter the next night nudge one to her particular reverend, trusting that the gossip so delivered would go no further than his jesuit's cloth, yet it was this overspoiled priest, Mr Browne, who was overheard to pianissime a slightly varied version and hushly pierce the rubiend aurellum of a layteacher of rural science during a priestly flutter at the hippic runfields.

I am currently writing a book about *Finnegans Wake*, and so getting to the underlying meaning of sentences is of absolute importance.

In order to understand the long sentences, it is imperative to determine what the essence is, as shown in the above example.

So as a first step in breaking down the complexity of Joyce's work, I decided I would try writing some ixml [ixml] to break down the sentences into their structural form. I quickly discovered that each chapter has its own peculiarities, so I am writing a different grammar per chapter, in order to keep it simpler.

The First Pass

So I started off as simple as can be with the following grammar:

```
chapter: paragraph+. {a chapter is one or more paragraphs}
paragraph: line+, #a. {a paragraph is one or lines,
                      followed by a blank line}
line: ~[#a]+, #a. {a line is characters (except end of line),
                  followed by end-of-line}
```

This failed because (obviously, once I had thought about it) the last paragraph is not followed by a blank line. A blank line *separates* paragraphs:

```
chapter: paragraph++#a.
paragraph: line+.
line: ~[#a]+, #a.
```

This now produced a first, very basic output (not shown here). But we are not interested in lines, but in the internal structure of paragraphs. Let's try:

```
chapter: paragraph++#a. {a chapter is one or more paragraphs}
paragraph: sentence+. {a paragraph is one or more sentences}
sentence: phrase++punctuation, ".". {A sentence is one or more phrases,
                                     separated by some punctuation,
```

phrase: word++ " ".	terminated with a point} {a phrase is one or more words, separated by spaces}
word: [L]+.	{a word is one or more letters}
punctuation: [",;:"].	{phrases are separated by these}

This failed on the very first line of the chapter:

```
Now concerning the genesis of Harold or Humphrey Chimpden's
**** Character: "'" (#2019).
```

Ah, a word consists of more than letters. Fix that:

```
word: [L; "'"]+.
```

and it immediately failed on the same line:

```
**** Parsing failed at line 1, position 58
Now concerning the genesis of Harold or Humphrey Chimpden's
**** Character: "(#A).
```

Of course, words are not only separated by spaces, but sometimes by newlines.

Change

```
phrase: word++ " ".
```

to

```
phrase: word++s.
s: [" "; #a].
```

At least we now got to line 2:

```
occupational agnomen, the best authenticated version has it that it
**** Character: " " (#20).
```

Oh yes, punctuation can also be followed by space...

```
punctuation: [",;:", s*.
```

Now we get to line 3:

```
was this way. We are told how in the beginning it came to pass that
**** Character: " " (#20).
```

Ah, full-stops can be followed by space as well:

```
sentence: phrase++punctuation, ".", s*.
```

This gets us to line 17!

```
seldomer than an earwigger! Comes the question are these the facts of
**** Character: "!" (#21).
```

Of course, sentences can also end with "!". Let's also add "?", just in case:

```
sentence: phrase++punctuation, [".!?"], s*.
```

At line 62 we discover another character that can appear in a word:

```

Ides-of-April morning (the anniversary, as it fell out, of his first
^
**** Character: "-" (#2D).

```

Fix that:

```
word: [L; "'-"]+.
```

And on the same line, we come to our first structuring problem:

```

Ides-of-April morning (the anniversary, as it fell out, of his first
^
**** Character: "(" (#28).

```

Nested phrases! So where do we put this in the structure? I'm going to experiment, and put it at the level of a word. First separate the definition of phrases:

```

sentence: phrases, [".!?"], s*.
phrases: phrase++punctuation.

```

and add bracketed phrases:

```

word: [L; "'-"]+; bracketed.
bracketed: "(", phrases, ")".

```

This does well, and gets us to line 136, when we get this surprise:

```

fellow-me-lieder was first poured forth to an overflow meeting of all
^
**** Character: "-" (#2014).

```

The typesetters had used two different characters for hyphenated words! Fix that:

```
word: [L; "'--"]+; bracketed.
```

And we get to the end of the chapter! Hooray! The output starts:

```

<chapter ixml:state="ambiguous" xmlns:ixml="http://invisiblexml.org/NS">
  <paragraph>
    <sentence>
      <phrases>
        <phrase>
          <word>Now</word>
          <s> </s>
          <word>concerning</word>
          <s> </s>
          <word>the</word>
          <s> </s>
          <word>genesis</word>
          <s> </s>
          <word>of</word>
          <s> </s>
          <word>Harold</word>
          <s> </s>
          <word>or</word>
          <s> </s>
          <word>Humphrey</word>
          <s> </s>
          <word>Chimpden's</word>
          <s> </s>
          <word>occupational</word>
          <s> </s>
          <word>agnomen</word>
        </phrase>
      <punctuation>,
      <s> </s>
    </sentence>
  </paragraph>
</chapter>

```

Cleaning up the Output

We'll look at the ambiguity in a minute, but first to get rid of elements we don't need, by adding "-" before some rules. This produces a better looking result:

```
<chapter ixml:state="ambiguous" xmlns:ixml="http://invisiblexml.org/NS">
  <paragraph>
    <sentence>
      <phrase>Now concerning the genesis of Harold or Humphrey Chimpden's
occupational agnomen</phrase>,
      <phrase>the best authenticated version has it that it
was this way</phrase>. </sentence>
    <sentence>
      <phrase>We are told how in the beginning it came to pass that
the grand old gardener was saving daylight under his redwoodtree one
sultry sabbath afternoon</phrase>,
      <phrase>when royalty was announced to have been
pleased to have halted itself on the highroad</phrase>. </sentence>
```

What is also obvious is that the newline characters in the input are visible in the output. On the other hand we want to keep the spaces. So we could change

```
s: [" "; #a].
```

to

```
s: " "; -#a.
```

but then words separated by a newline will run together; so we replace newlines with spaces:

```
s: " "; -#a, +" ".
```

An alternative is to delete all whitespace, and replace it with a single space character:

```
s: -[" "; #a], +" ".
```

Now we get

```
<chapter ixml:state="ambiguous" xmlns:ixml="http://invisiblexml.org/NS">
  <paragraph>
    <sentence>
      <phrase>Now concerning the genesis of Harold or Humphrey Chimpden's
occupational agnomen</phrase>,
      <phrase>the best authenticated version has it that it was
this way</phrase>. </sentence>
    <sentence>
      <phrase>We are told how in the beginning it came to pass that
the grand old gardener was saving daylight under his
redwoodtree one sultry sabbath afternoon</phrase>,
      <phrase>when royalty was announced to have been pleased to have
halted itself on the highroad</phrase>. </sentence>
```

If you're interested in how the bracketed phrases look, here's an example:

```
<sentence>
  <phrase>They tell the story how one happygogusty Ides-of-April morning
    <bracketed>(
      <phrase>the anniversary</phrase>,
      <phrase>as it fell out</phrase>,
      <phrase>of his first assumption of his mirthday suit</phrase>)
    </bracketed> ages and ages after the alleged misdemeanour
when the tried friend of all creation was billowing across
the wide expanse of our greatest park</phrase>,
  <phrase>he met a cad with a pipe</phrase>. </sentence>
```

Dealing with Ambiguity

The output claims there are ten different interpretations of the chapter, so let's have a look why.

```
The input from line.pos 1.1 to 148.1 can be interpreted as 'paragraph++a'
in 10 different ways:
1: paragraph++a[1.1]: paragraph[:20.1] #a[:21.1] paragraph++a[:148.1]
2: paragraph++a[1.1]: paragraph[:31.1] #a[:32.1] paragraph++a[:148.1]
3: paragraph++a[1.1]: paragraph[:87.1] #a[:88.1] paragraph++a[:148.1]
4: paragraph++a[1.1]: paragraph[:94.48] #a[:95.1] paragraph++a[:148.1]
5: paragraph++a[1.1]: paragraph[:102.29] #a[:103.1] paragraph++a[:148.1]
6: paragraph++a[1.1]: paragraph[:109.1] #a[:110.1] paragraph++a[:148.1]
7: paragraph++a[1.1]: paragraph[:121.1] #a[:122.1] paragraph++a[:148.1]
8: paragraph++a[1.1]: paragraph[:139.1] #a[:140.1] paragraph++a[:148.1]
9: paragraph++a[1.1]: paragraph[:145.1] #a[:146.1] paragraph++a[:148.1]
10: paragraph++a[1.1]: paragraph[:148.1]
```

and if we look at the input, we find that each of the lines it mentions (20, 31, 87, etc), are paragraph breaks. The problem is that we have allowed sentences to be separated by more than one space, which also matches the extra newline after a paragraph.

```
sentence: phrases, ["!?"], s*.
```

So let's delete that *. This exposes a new problem:

```
Haromphrey bear the sigla H.C.E. and while he was only and long and
                        ^
**** Character: "C" (#43).
```

Full-stops are not only used to separate sentences! "H.C.E." clearly looks like it could end a sentence, because it ends with a full-stop and a space. We'll have to treat it as a special kind of word:

```
-word: [L; "'--" ]+; bracketed; initialism.
initialism: ([Lu], ".")+.
```

(Lu matches any upper-case letter).

This reveals another ambiguity, only this time, it's a real one, and not a mistake in the grammar:

```
To anyone who knew and loved the christlikeness of the big cleanminded
giant H. C. Earwicker throughout his excellency long vicefreegal existence
```

We can see as humans that this is not ambiguous, but consider this sentence:

```
There are people who claim sentences
never end with a capital H. C. Earwicker
however, in his seminal paper "Sentences
that end with a capital H", proves otherwise.
```

What we get is

```
<sentence>
  <phrase>To anyone who knew and loved the christlikeness of
    the big cleanminded giant H</phrase>. </sentence>
<sentence>
  <phrase>C</phrase>. </sentence>
<sentence>
  <phrase>Earwicker throughout his excellency long vicefreegal existence ...
```

To tell you the truth, at this point I cheated. I deleted the first space in "H. C. Earwicker", and the chapter was parsed to completion without ambiguity.

Another Chapter

I presented the development of chapter 2 above, since it is fairly simply structured. The sentence presented at the beginning of the paper on the other hand is from chapter 3, which is more complicated. So to end, I

will simply show the current ixml for chapter 3, and the result of the parsing of the example sentence. You will see that I have handled a number of aspects, such as punctuation, differently.

```

chapter: paragraph++(-#a, -#a), -#a.
paragraph: sentence++s; pagenumber.
pagenumber: ["0"-"9"]+.
-sentence: question; exclamation; statement.
question: phrases, "?".
exclamation: phrases, "!".
statement: phrases, ".".
-phrases: punctuated-phrase*, phrase.
punctuated-phrase>phrase: -phrase, ["?!"]?, punc.
-s: -" "+, -#a?; -#a.
phrase: word++(s, +" ").
-word: bit++("-", (-#a, +"?")?); bracketed.
-bit: ([L; "0"-"9"; #2019]; ".", ~[" "; #a])+.
bracketed: s?, -"(", (phrases; -paragraph), -")";
s?, -"-", s?, phrases, -"-".
@punc: [",;:"], s.

```

which applied to the example sentence produces a structure like this:

```

<statement>
  <phrase punc=', '>Our cad's bit of strife
    <bracketed>
      <phrase>knee Bareniece Maxwellton</phrase>
    </bracketed> with a quick ear for spittoons
    <bracketed>
      <phrase>as the aftertale hath it</phrase>
    </bracketed>
    glaned up as usual with dumbestic husbandry
    <bracketed>
      <exclamation>
        <phrase punc=', '>no persicks and armelians for thee
        </phrase>
        <phrase>Pome-?ranzia</phrase>!
      </exclamation>
    </bracketed> but
  </phrase>
  <phrase punc=', '>slipping the clav in her claw</phrase>
  <phrase punc=', '>broke of the matter among a hundred
    and eleven others in her usual curtsey
    <bracketed>
      <exclamation>
        <phrase punc=', '>how faint these first vhespers womanly are
        </phrase>
        <phrase punc=', '>a secret pispigliando</phrase>
        <phrase>amad the lavurdy den of their manfolker
        </phrase>!
      </exclamation>
    </bracketed>
    the next night nudge one as was
    Hegesippus over a hup a ' chee
  </phrase>
  <phrase punc=', '>her eys dry and small and speech
    thicklish because he appeared a funny colour like
    he couldn't stood they old hens no longer
  </phrase>
  <phrase punc=', '>to her particular reverend</phrase>
  <phrase punc=', '>the director</phrase>
  <phrase punc=', '>whom she had been meaning in her mind
    primarily to speak with
    <bracketed>
      <exclamation>
        <phrase punc=', '>hosch</phrase>
        <phrase>intra</phrase>!
      </exclamation>
      <exclamation>
        <phrase>jist a timblespoon</phrase>!
      </exclamation>
    </bracketed>
    trusting
  </phrase>
  <phrase punc=', '>between cuppled lips and annie lawrie promises
    <bracketed>
      <exclamation>
        <phrase>mighshe never have Esnekerry pudden come

```

```

        Hunanov for her pecklapitschens
    </phrase>
</exclamation>
</bracketed>
    that the gossip so delivered in his epistolear
</phrase>
<phrase>buried teatoastally in their Irish stew would go no
    further than his jesuit's cloth
</phrase>
</statement>

```

The only thing I will explain here is this:

```
<phrase>Pome-?ranzia</phrase>!
```

If a word is hyphenated over the end of a line (as this word was), you can't tell if the hyphen is meant to be part of the word, or is only there to signal a word split over two lines. So I add a question mark after such a hyphen (since ends of line are deleted), to make it clear that this is a special type of hyphen.

Conclusion

This paper is intended to give insight into the processes that an ixml grammar author can go through while trying to describe a document whose structure is not yet completely obvious. Writing grammars is a learned skill, that needs experience to gain fluency. In particular, learning to deal with ambiguity is difficult, because what is ambiguous to the computer doesn't always appear ambiguous to the human eye. However, once learned, the ability to parse large texts can help simplify enormously the automatic processing of large documents.

References

- [Kells] Anonymous, *Book of Kells*, Wikipedia, 2025, https://en.wikipedia.org/wiki/Book_of_Kells
- [fw] James Joyce, *Finnegans Wake*, Faber and Faber, 1939
- [ixml] Steven Pemberton (ed.), *Invisible XML Specification*, Invisible XML Organisation, 2022, <https://invisiblexml.org/1.0/>

Author's keywords for this paper: ixml; parsing; Finnegans Wake

Steven Pemberton

Researcher

CWI, Amsterdam

<steven.pemberton@cwi.nl>

Steven Pemberton is a researcher affiliated with CWI, Amsterdam. His research is in interaction, and how the underlying software architecture can support users.

He co-designed the ABC programming language that formed the basis for Python and was one of the first handful of people on the open internet in Europe, when the CWI set it up in 1988. Involved with the Web from the beginning, he organised two workshops at the first Web Conference in 1994. For the best part of a decade he chaired the W3C HTML working group, and has co-authored many web standards, including HTML, XHTML, CSS, XForms and RDFa. He now chairs the W3C XForms and Invisible Markup groups. In 2022, ACM SIGCHI awarded him the Lifetime Practice Award. More details at <http://www.cwi.nl/~steven>

Balisage Series on Markup Technologies