



# Element classification

*A bottom up perspective*

Diederik Gerth van Wijk  
dg@doxatrix.nl

Declarative Amsterdam 2022

# This talk top down



- Top down and bottom up
- Content models: order, cardinality
- Golden rules of content modeling
- Detour: the missing parent
- Sequence paradoxes
- BUCL up
- Round up

# Top down and bottom up



- XML trees are like family trees
- The root is the top
- Unlike people, elements have only one parent
- Element **types** however may be used in multiple element type definitions



# Content models



- Element type definitions only state the possible children: top down
- No control over parents, siblings or how the element may be used
- Is that a pity?
- Does the way an element type may be used say something about that type?

# Top down classification



- Element types are /have
  - empty
  - mixed
  - element content, with children in
    - a repeatable OR group
    - a non repeatable OR group which is in a SEQ group
    - a SEQ group

# Bottom up classification



- Take content models for *book*:  
(*title*, *author*+, (*sect*+ | *chap*+))  
and *sect*: (*para*|*table*|*list*)+
- An element is:
  - the root element (*book*)
  - in a sequence (*title*, *author*)
  - in an OR group (*sect*, *chap*)
  - in a repeatable OR group (*para*)
  - in mixed content, like *emph* in  
(#PCDATA | *emph* | *anch*)\*

# Combined classification



Occurs in:	Mixed content	Sequence group	Non repeatable OR group	Repeatable OR group	Root
Has:					
Empty element	<code>&lt;anchor/&gt;</code>	<code>&lt;keyword/&gt;</code>			
Mixed content	<code>&lt;emph&gt;</code>	<code>&lt;title&gt;</code>	<code>&lt;author&gt;</code>	<code>&lt;para&gt;</code>	
Element content	<code>&lt;formula&gt;</code>	<code>&lt;header&gt;</code>	<code>&lt;section&gt;</code>	<code>&lt;table&gt;</code>	<code>&lt;book&gt;</code>

# Golden rules



- Many language owners have rules like:
  - An element should occur only once in a content model.
  - Sequence groups should not be optional or repeatable.
  - Mixed content may not contain element content.
  - Element types should occur only in mixed or in element content.



# Element content in mixed?



- In simple text (*title*, *emph*)?  
Usually not.
- In complex text (*paragraph*)?  
Maybe:
  - table
  - formula
  - figure
  - list
  - quotation
  - note

# In or between *para* (1)



- A *table* or *figure* may be *rendered* in a paragraph
- Not part of a sentence
- Encode *between* paragraphs and put an anchor for placement

## EQUILIBRIUM IN THE LABOR MARKET

The labor demand curves for both competitive and monopolistic firms with fixed capital stock were developed in Chapter 6. The firms are assumed to have production functions  $y = y(N; \bar{K})$  with marginal product of labor  $MPL = \partial y / \partial N$ . Then the competitive firm's labor demand (or wage offer) curve is given by

$$(4) \quad W = P \cdot \frac{\partial y}{\partial N},$$

which sets the real wage  $w$  equal to the  $MPL$ . The monopolist's labor demand curve is given by

$$(5) \quad W = P \cdot \left(1 + \frac{1}{e}\right) \cdot \frac{\partial y}{\partial N},$$

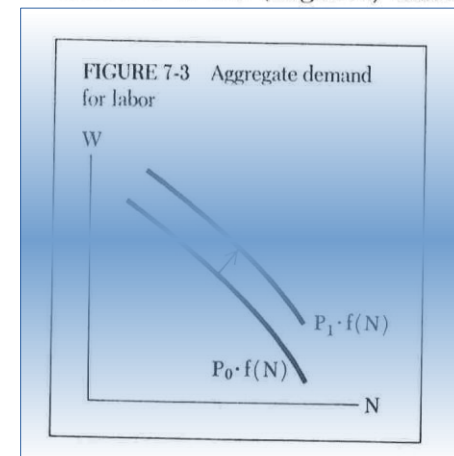
where  $e$  is the (negative) elasticity of the demand curve for the monopolist's output.

In an economy with a mixture of competitive and monopolistic firms, the aggregate labor-demand curve will look like an "average" of (4) and (5):

$$(6) \quad W = P \cdot f(N); \quad f' < 0.$$

Here  $f(N)$  is related to the economy's aggregate  $MPL$  function and has a negative slope. When we plot this aggregate demand function in the  $W, N$  space of Figure 7-3, we see that an increase in the price level from  $P_0$  to  $P_1$  shifts the labor demand curve

up by raising the value of the marginal product of labor. This contrasts with the behavior of labor supply in this money wage model; a price change does not shift the labor supply curve in the  $W, N$  space.



# In or between *para* (2)



- Formulae, lists and quotations *can* be part of a sentence:
- Encode *in* the paragraph!

## EQUILIBRIUM IN THE LABOR MARKET

The labor demand curves for both competitive and monopolistic firms with fixed capital stock were developed in Chapter 6. The firms are assumed to have production functions  $y = y(N, \bar{K})$  with marginal product of labor  $MPL = \partial y / \partial N$ . Then the competitive firm's labor demand (or wage offer) curve is given by

$$(4) \quad W = P \cdot \frac{\partial y}{\partial N}$$

which sets the real wage  $w$  equal to the  $MPL$ . The monopolist's labor demand curve is given by

$$(5) \quad W = P \cdot \left(1 + \frac{1}{e}\right) \frac{\partial y}{\partial N}$$

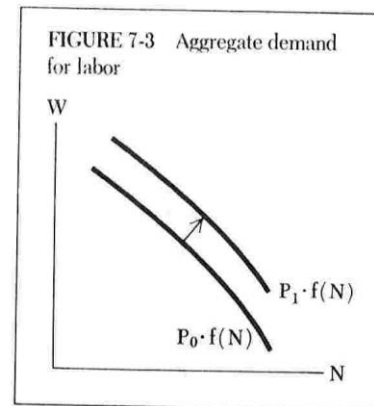
where  $e$  is the (negative) elasticity of the demand curve for the monopolist's output.

In an economy with a mixture of competitive and monopolistic firms, the aggregate labor demand curve will look like an "average" of (4) and (5):

$$(6) \quad W = P \cdot f(N); \quad f' < 0.$$

Here  $f(N)$  is related to the economy's aggregate  $MPL$  function and has a negative slope. When we plot this aggregate demand function in the  $W, N$  space of Figure 7-3, we see that an increase in the price level from  $P_0$  to  $P_1$  shifts the labor demand curve

up by raising the value of the marginal product of labor. This contrasts with the behavior of labor supply in this money wage model; a price change does not shift the labor supply curve in the  $W, N$  space.



# In or between *para* (3)



- The sentence may continue after the list!
- A list item can, while being part of a sentence, contain a second sentence!

Art. 19  
Vorderingen bij derden

1. Een derde die:

- loon;
- uitkeringen op grond van sociale zekerheidswetten, uitgezonderd kinderbijslag onder welke benaming ook;
- pensioen en lijfrente;
- uitkeringen uit levens-, invaliditeits-, ongevallen- of ziekengeldverzekering;
- uitkeringen tot levensonderhoud, verschuldigd krachtens boek I van het Burgerlijk Wetboek, of tot vergoeding van schade door het derven van levensonderhoud;
- uitkeringen of buitengewone pensioenen op grond van een wettelijke regeling voor oorlogsgetroffenen of voor degenen die hun dienstplicht of vervangende dienstplicht vervullen;
- bedragen, toegekend krachtens artikel 8, derde lid, van het koninklijke besluit van 17 juli 1944, *Stb.* E 51, betreffende de arbeidsbemiddeling en de scholing, herscholing en omscholing;

verschuldigd is aan een belastingschuldige, is op vordering van de

Art. 143  
Bewijs van aanvangswaarde van stamvermogen

De aanvangswaarde van de tot het stamvermogen behorende goederen wordt als volgt bewezen:

- wat betreft de bij het aangaan van het deelgenootschap aanwezige goederen, uitsluitend door de in het vorige artikel bedoelde staat. Ontbreekt een goed op die staat of is de waarde daarvan niet daarbij vermeld, dan komt de waarde van dat goed niet voor de berekening van de waarde van het stamvermogen van de betrokken echtgenoot in aanmerking;
- wat betreft de door erfopvolging, making of gift verkregen goederen, door de aangifte volgens welke het recht van successie, schenking of overgang is gegeven; is een aanslag niet overeenkomstig de aangifte geschied, dan wordt die aanslag aan het bewijs ten grondslag gelegd. Bij gebreke van een aangifte en een aanslag, kan het bewijs door alle middelen rechtens worden geleverd. (Ontw. 1.8.2.11; Overg.w. 13 lid 3; Succ.w. 36 e.v.)

# Element content in mixed?



- There is a use case
- Coding lists, formulae and quotations *in* a paragraph:
  - better classification of paragraphs by relative word frequency
  - consistent layout (indent, white lines)

# Detour: a missing element



- Most DTDs lack what *should* be the 2<sup>nd</sup> most used element:
- an element that groups some paragraphs:
- a *section* without *title* (top down)
- or a *paragroup* (bottom up)

truffling pig and presently gave a cry of triumph. He had disclosed one of those deposits of rubbish which are dear to the private soldier's sense of order: the head of a broom, the lid of a stove, a bucket rusted through, a sock, a loaf of bread, lay under the dock and nettle among cigarette packets and empty tins.

'Look at that,' said the commanding officer. 'Fine impression that gives to the regiment taking over from us.

'That's bad,' I said.

'It's a disgrace. See everything there is burned before you leave camp.'

'Very good, sir. Sergeant-major, send over to the carrier-platoon and tell Captain Brown that the C.O. wants this ditch cleared up.'

I wondered whether the colonel would take this rebuff; so did he. He stood a moment irresolutely prodding the muck in the ditch, then he turned on his heel and strode away.

'You shouldn't do it, sir, ' said the sergeant-major, who had been my guide and prop since I joined the company. 'You shouldn't really.'

'That wasn't our rubbish.'

'Maybe not, sir, but you know how it is. If you get on the wrong side of senior officers they take it out of you other ways.'

As we marched past the madhouse, two or three elderly inmates gibbered and mouthed politely behind the railings.

'Cheeroh, chum, we'll be seeing you'; 'We shan't be long now'; 'Keep smiling till we meet again', the men called to them.

I was marching with Hooper at, the head of the leading platoon.

'I say, any idea where we're off to? '

'None.'

'Do you think it's the real thing?'

'No.'

'Just a flap?'

'Yes.'

'Everyone's been saying we're for it. I don't know what to think really. Seems so silly somehow, all this drill and training if we never go into action.'

'I shouldn't worry. There'll be plenty for everyone in time.'

'Oh, I don't want much you know. Just enough to say I've been in it.'

A train of antiquated coaches was waiting for us at the siding; an R.T.O. was in charge; a fatigue party was loading the last of the kit-bags from the trucks to the luggage vans. In half an hour we were ready to start and in an hour we started.

# Top down content models



- *para*:  
(#PCDATA | emph | anch | list |  
quote | formula | noteref |  
tableref | figureref)\*
- *paragroup*:  
(para | note | table | figure)+

# Sequence groups



- Element content:  
a sequence of OR groups
- A very generic model for novels:  
(title, author+, paragroup\*, section\*,  
chapter\*, part\*)
- Or a stricter submodel:  
(title, author+, (paragroup+ |  
                  section+ | chapter+ | part+))
- (it is a *submodel* because the set of  
valid instances is a subset of the set  
of valid instances of the supermodel)



# First sequence paradox



- Take the stricter model:  
(title, author+, (paragroup+ |  
                  section+ | chapter+ | part+))
- One title, one or more authors, one or more sections etc.
- A sequence group is the only way to specify cardinality
- A sequence group is not about the sequence, it is about the cardinality

# Second sequence paradox



- Valid instances of this model:  
(title, author+, chapter+)  
contain exactly the same information as  
valid instances of  
(author+, title, chapter+)  
or even  
(chapter+, author+, title)
- A simple XSLT can transform a “wrong”  
sequence in the right one
- **If the order according to the DTD  
matters, the order in the instances  
doesn't matter**

# BUC L up (1)



- Appearing in sequence groups or in mixed content says something about an element:
  - Elements that appear in element content are typically block elements in CSS
  - Elements that appear in element content are typically word boundaries for word searching or counting

# BUC L up (2)



- A *title* should be unique: it is a property of the element type
- The element type definition should specify such properties
- Need for a  
**Class Oriented Element Type  
Definition Language**  
or  
**Bottom Up Constraint Language**

# The potential of BUCL



- Take the book models:
  1. (title, author+, paragroup\*, section\*, chapter\*, part\*)
  2. (title, author+, (paragroup+ | section+ | chapter+ | part+))
- *book*: contains(\_headelements, \_bodyelements)
- *title*: is\_in(\_headelements)  
is(single, required)
- *author*: is\_in(\_headelements)  
is(repeatable, required)

# The potential of BUCL



- *paragraph*:  
  `is_in_class(_bodyelements)`  
  `contains((para|...|figure)+)`
- *section*:  
  `is_in_class(_bodyelements)`
- *chapter*:  
  `is_in_class(_bodyelements)`
- *part*:  
  `is_in_class(_bodyelements)`

# The potential of BUCL



- *\_bodyelements*:  
contains(title,  
\_bodyelements(sequence, optional,  
repeatable, nonrecursive))
- creates  
*book*: (title, author+, paragroup\*,  
section\*, chapter\*, part\*)  
*part*: (title, paragroup\*, section\*,  
chapter\*)  
*chapter*: (title, paragroup\*, section\*)  
*section*: (title, paragroup\*)

# The potential of BUCL



- *\_bodyelements*:  
contains(title,  
\_bodyelements(**choice**, **required**,  
repeatable, nonrecursive))
- creates  
*book*: (title, author+, (paragroup+ |  
section+ | chapter+ | part+))  
*part*: (title, (paragroup+ | section+ |  
chapter+))  
*chapter*:  
(title, (paragroup+ | section+))  
*section*: (title, paragroup+)



# Round up



- Defining how an element should be used in content models leads to consistency (cardinality, order)
- Whether an element may be used in element or in mixed content matters (CSS, word boundaries)
- Element classes are a solution
- BUCL could help, doesn't exist yet

# Questions?



- Ask me now
- Mail me:  
dg@doxatrix.nl
- App, text or call:  
+31 622 27 81 55