合一算法
Unification

詹卫东
zwd@pku.edu.cn

冯志伟、孙乐译《自然语言处理综论》电子工业出版社2005年版。第11章。
Daniel Jurafsky & James H. Martin, 2000, Speech and Language Processing, Pearson Education, Inc., Prentice Hall.

例1

$$F1=\begin{pmatrix} \text{NUMBER:} & \text{SG} \end{pmatrix} \qquad F2=\begin{pmatrix} \text{PERSON:} & 3 \end{pmatrix}$$

$$F1 \overline{\cup} F2 = \begin{pmatrix} \text{NUMBER:} & \text{SG} \\ \text{PERSON:} & 3 \end{pmatrix}$$

例2

$$F1=\begin{bmatrix} \text{NUMBER:} & \text{SG} \end{bmatrix} \qquad F2=\begin{bmatrix} \text{NUMBER:} & \text{PL} \end{bmatrix}$$

$$F1 \overline{\cup} F2 = \text{FAILURE}$$

例3

$$F1=\begin{pmatrix} \text{NUMBER:} & [\ ] \end{pmatrix} \qquad F2=\begin{pmatrix} \text{NUMBER:} & \text{PL} \end{pmatrix}$$

$$F1 \overline{\cup} F2 = \text{NUMBER: PL}$$

例4

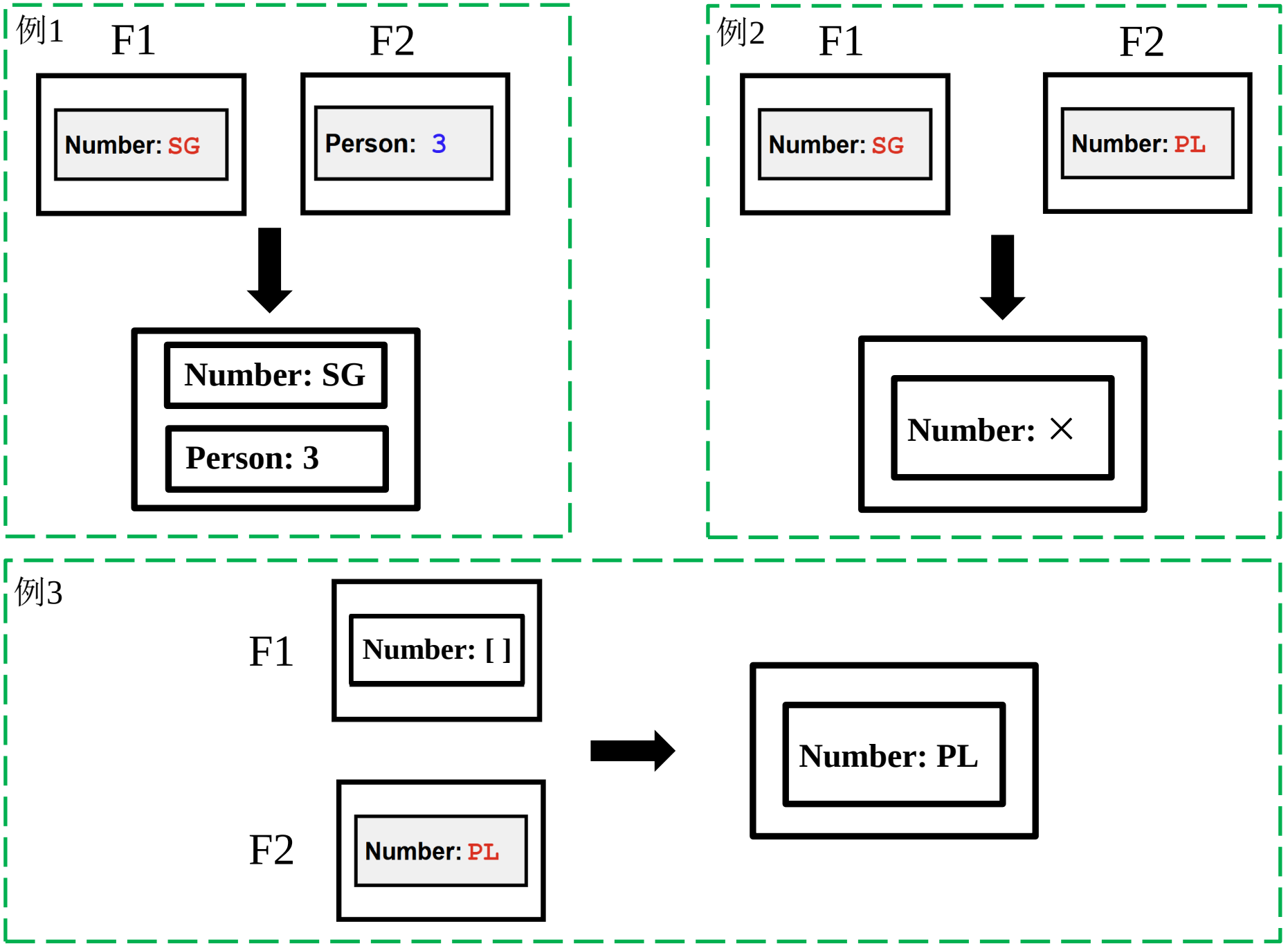
$$F1 = \begin{bmatrix} \text{Agree: } [\text{Number: SG}]\ ① \\ \text{Subject: } [\text{Agree: } ①] \end{bmatrix}$$

is happy

$$F2 = [\text{Subject: } [\text{Agree: } [\text{Person: 3}]]]$$

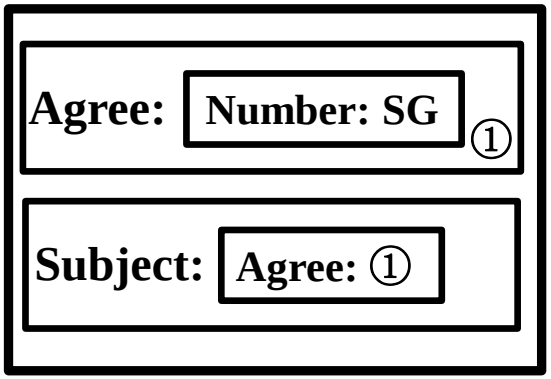
she

$$F1 \overline{\cup} F2 = \begin{bmatrix} \text{Agree: } \begin{bmatrix} \text{Number: SG} \\ \text{Person: 3} \end{bmatrix}\ ① \\ \text{Subject: } [\text{Agree: } ①] \end{bmatrix}$$

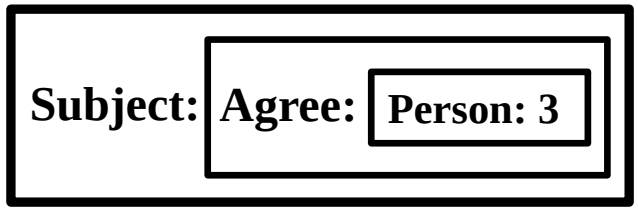

例1
F1
F2
Number: SG
Person: 3
Number: SG
Person: 3
例2
F1
F2
Number: SG
Number: PL
Number: ×
例3
F1
Number: [ ]
F2
Number: PL
Number: PL

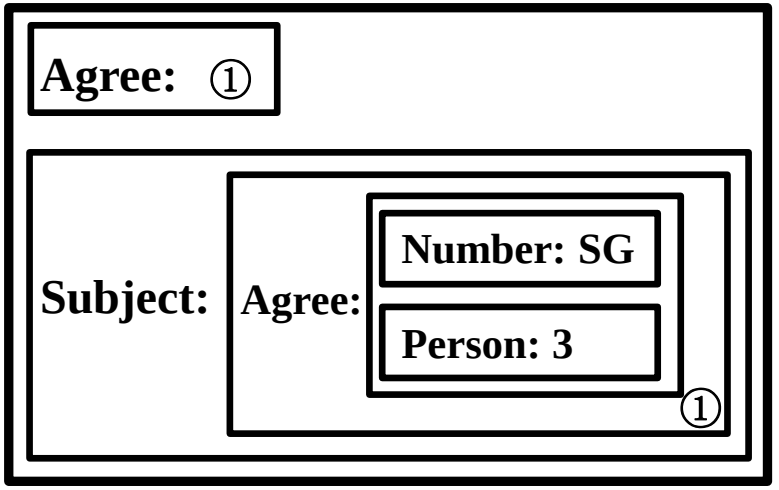
例4

F1

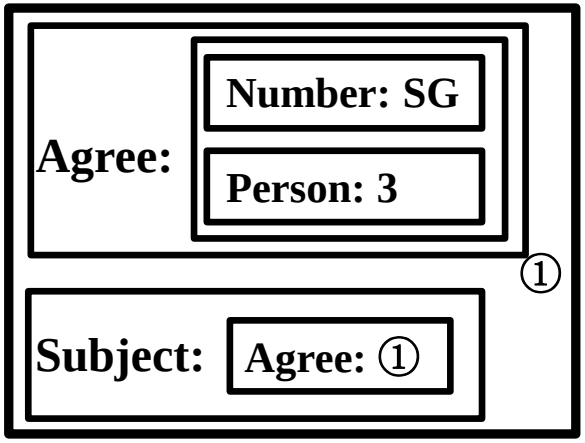

Agree:
Number: SG
①
Subject:
Agree: ①

F2

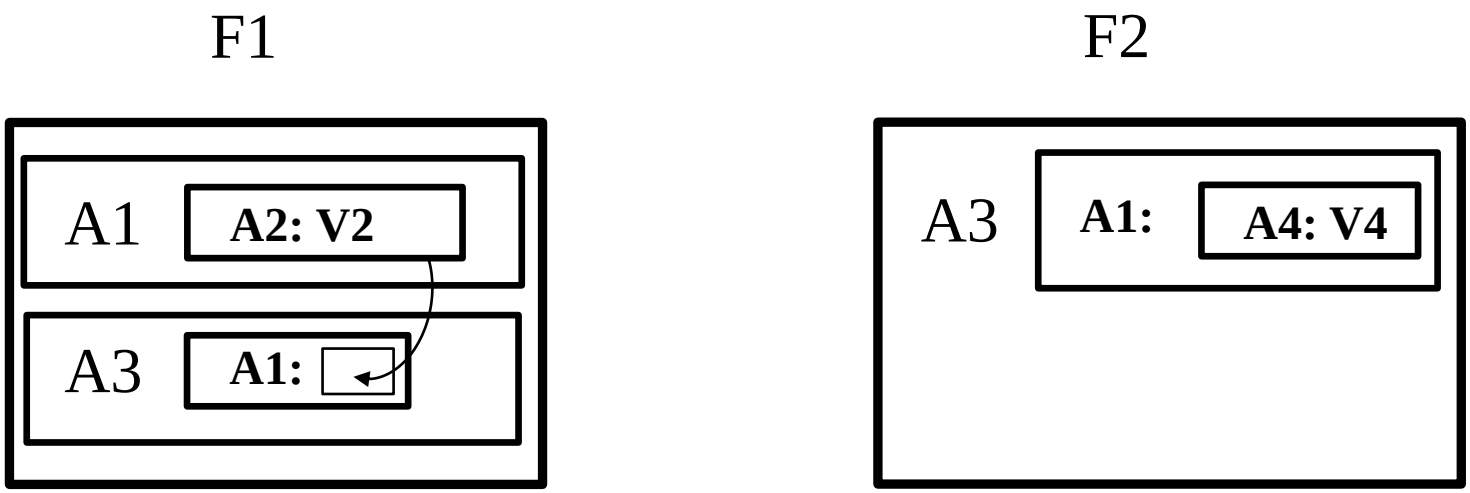

Subject:
Agree:
Person: 3


Agree: ①
Subject:
Agree:
Number: SG
Person: 3
①

=


Agree:
Number: SG
Person: 3
①
Subject:
Agree: ①

特征结构的合一运算



$F1 \overline{\cup} F2$　检查：F1和F2这两个盒子能否合并为一个？

(1) 盒子对齐

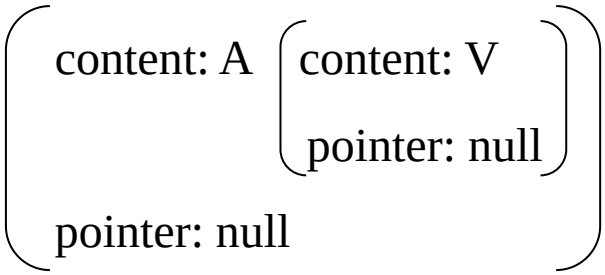
(2) 检查相同格子对物品要求是否冲突

特征结构的数据表示

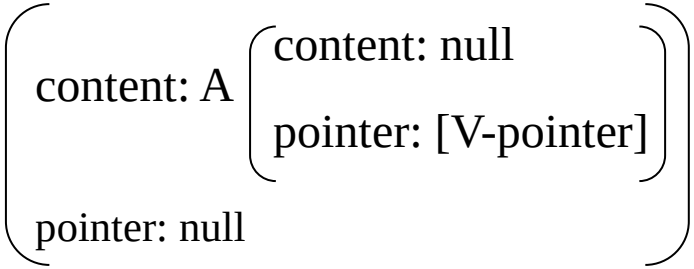
特征结构的形式改造（统一化）：　　新特征结构由两个字段content和pointer构成。

原特征结构F的特征值对 [A:V] 转为新特征结构时遵循以下规则：

如果V为原子或无共享特征结构，则新的特征结构为：



如果V为有共享值的特征结构，则新的特征结构为：



1. 便于算法统一描述（对齐）
2. Content 描述特征实质性的约束条件
3. Pointer 表示同指关系（共享特征值）

V-pointer指向V的共享值

通常用① ② ……表示。

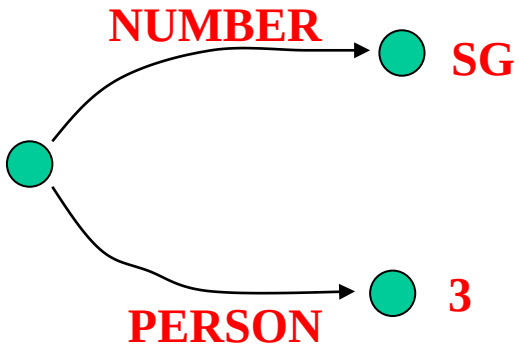
例1

$$F=\begin{bmatrix} \text{NUMBER} & \text{SG} \\ \text{PERSON} & 3 \end{bmatrix}$$

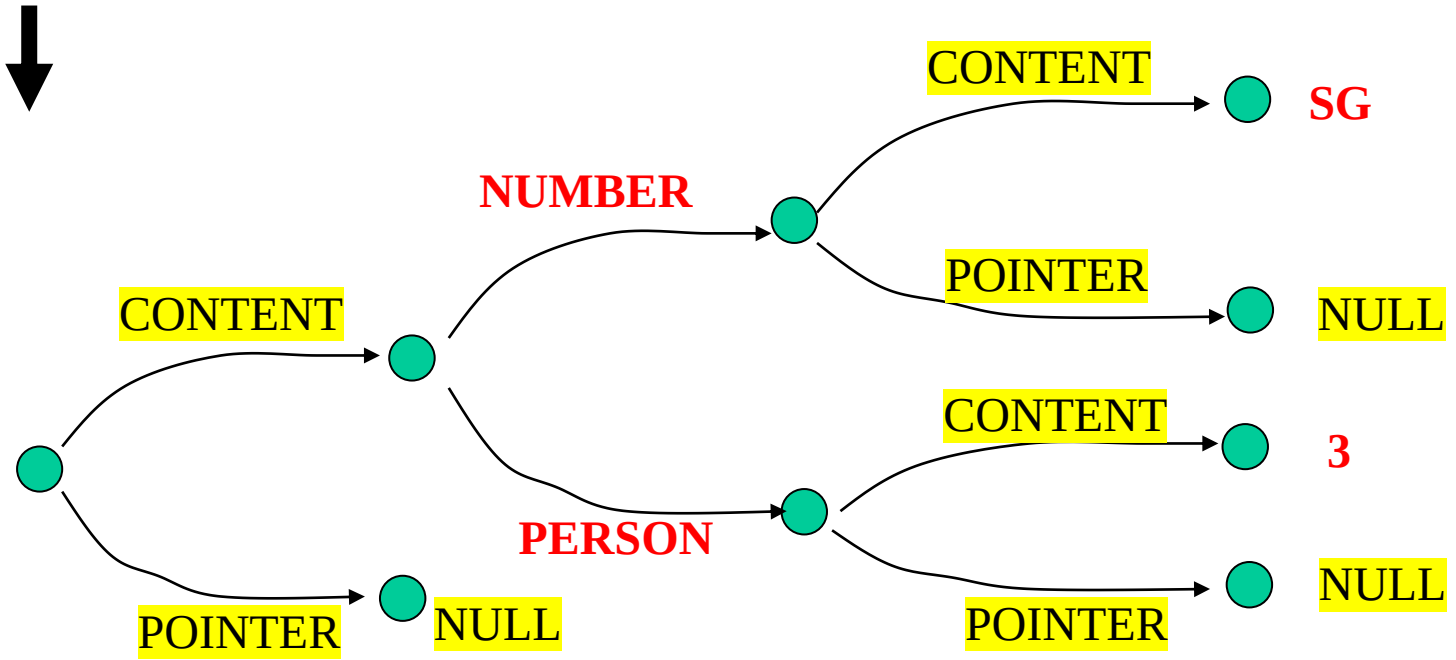
特征结构的统一形式

$$F=\begin{bmatrix} \text{CONTENT} & \begin{bmatrix} \text{NUMBER} & \begin{bmatrix} \text{CONTENT} & \text{SG} \\ \text{POINTER} & \text{NULL} \end{bmatrix} \\ \text{PERSON} & \begin{bmatrix} \text{CONTENT} & 3 \\ \text{POINTER} & \text{NULL} \end{bmatrix} \end{bmatrix} \\ \text{POINTER} & \text{NULL} \end{bmatrix}$$

例1

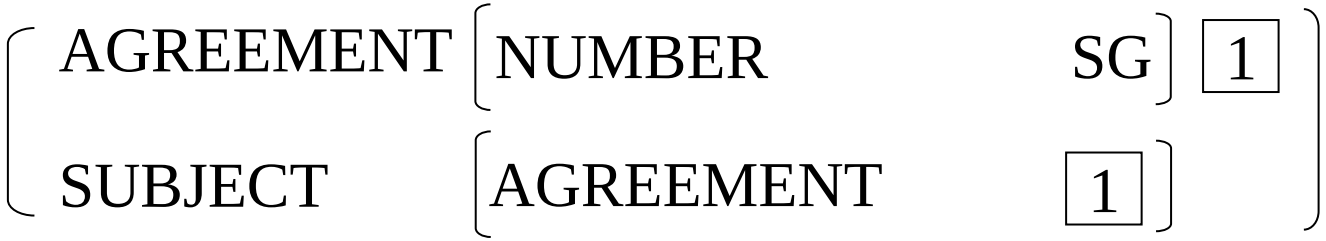


特征结构的统一形式

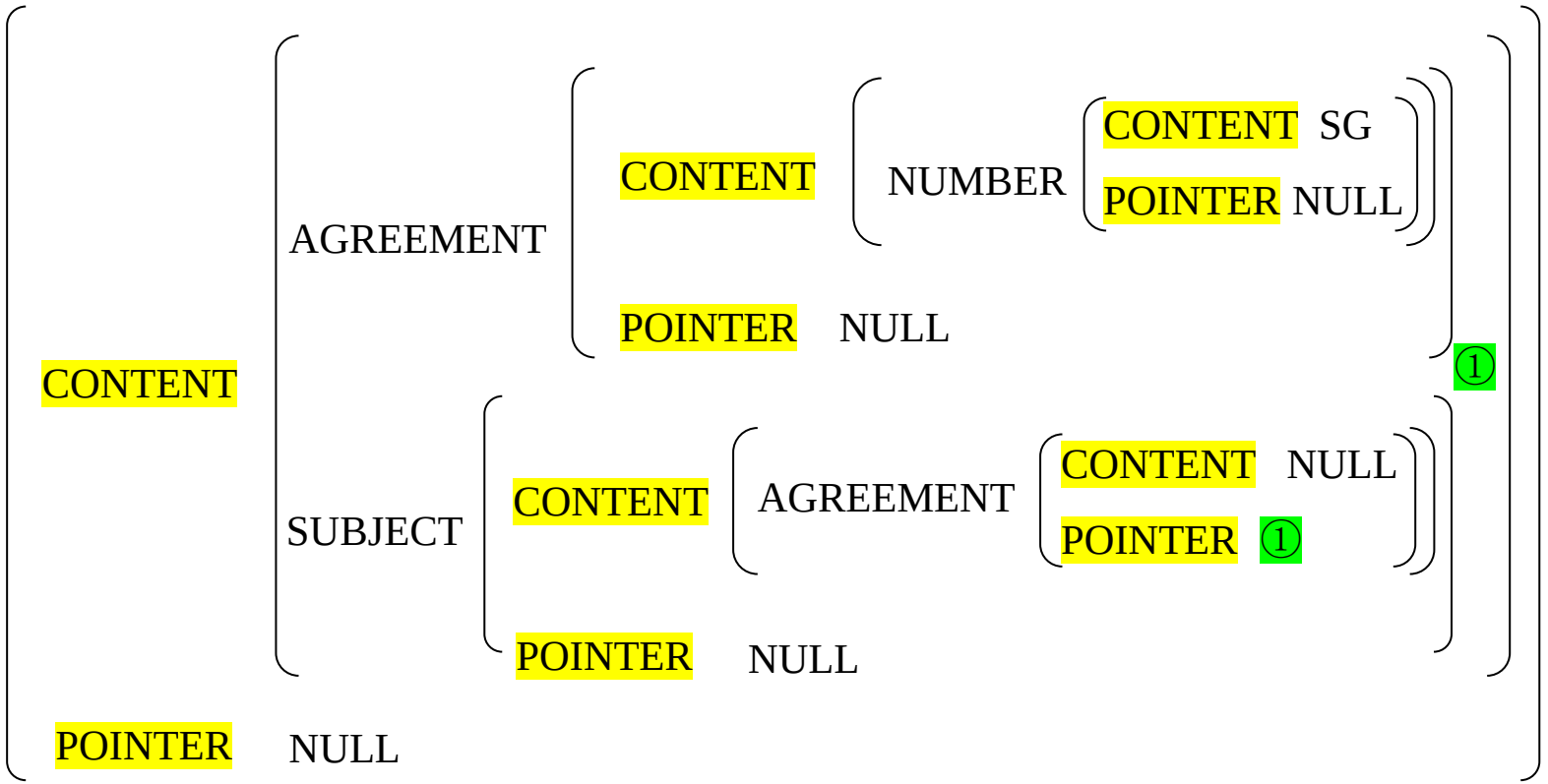


例4

F=

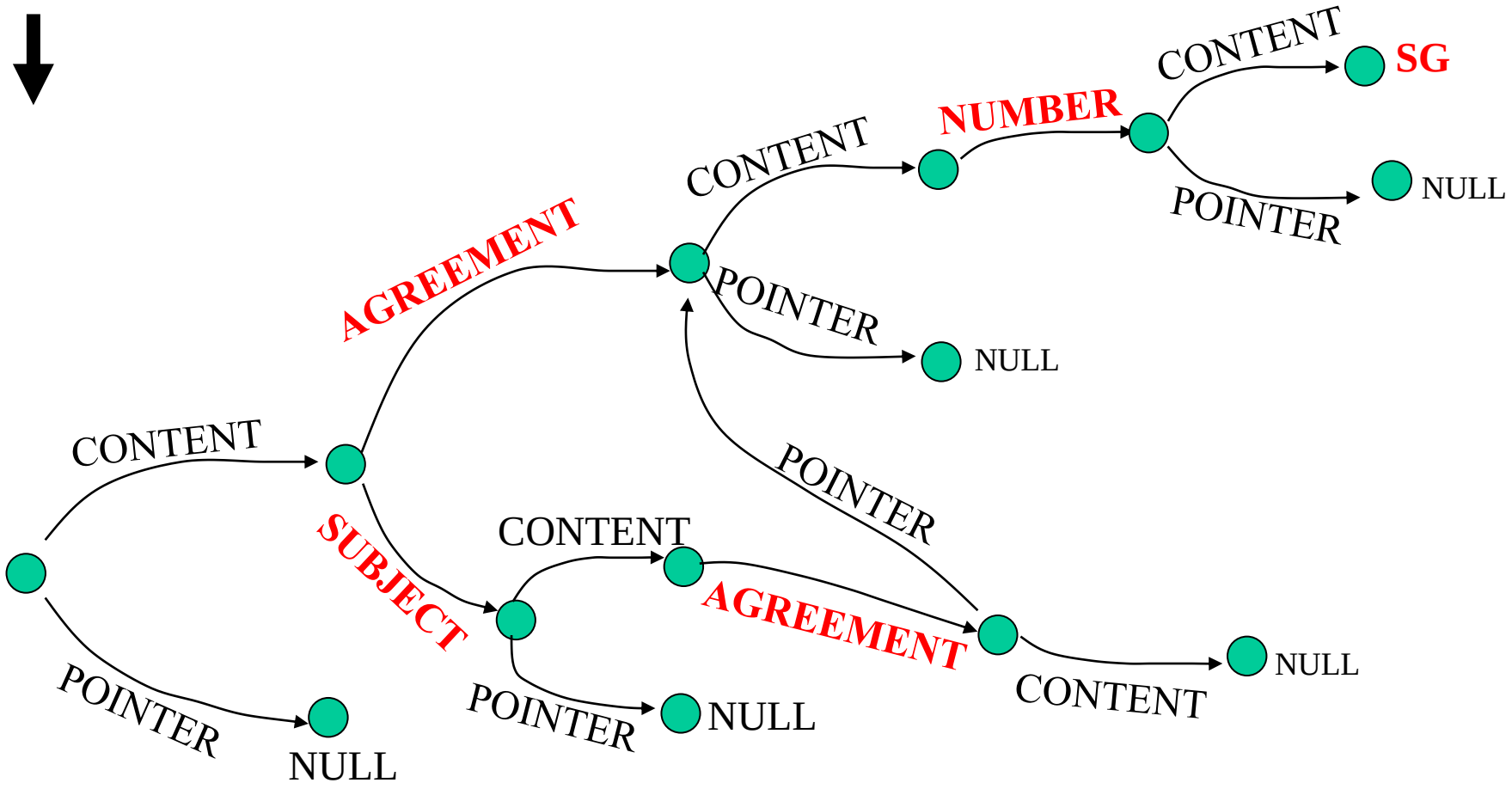

AGREEMENT
NUMBER
SG
1
SUBJECT
AGREEMENT
1

F=


CONTENT
AGREEMENT
CONTENT
NUMBER
CONTENT SG
POINTER NULL
POINTER NULL
①
SUBJECT
CONTENT
AGREEMENT
CONTENT NULL
POINTER ①
POINTER NULL
POINTER NULL

例4

$$F=\begin{bmatrix} \text{AGREEMENT} & \begin{bmatrix} \text{NUMBER} & \text{SG} & \boxed{1} \end{bmatrix} \\ \text{SUBJECT} & \begin{bmatrix} \text{AGREEMENT} & \boxed{1} \end{bmatrix} \end{bmatrix}$$



合一算法

分3种情况处理

```
function UNIFY(f1, f2)
1.    f1-real ← f1.content
2.    f2-real ← f2.content
3.    if f1-real is null then
4.        f1.pointer ← f2
5.        return f2
6.    else if f2-real is null then
7.        f2.pointer ← f1
8.        return f1
9.    else if f1-real and f2-real are identical then
10.       f1.pointer ← f2
11.       return f2
12.   else if both f1-real and f2-real are complex feature structures then
13.       f2.pointer ← f1
14.       for each feature in f2-real do
15.         f1-feature ← Find or create a feature corresponding to feature in f1-real
16.         if UNIFY(f1.feature.value, f2.feature.value) == failure then
17.             return failure
18.       return f1
19.   else return failure
```

(1) f1或f2是 NULL

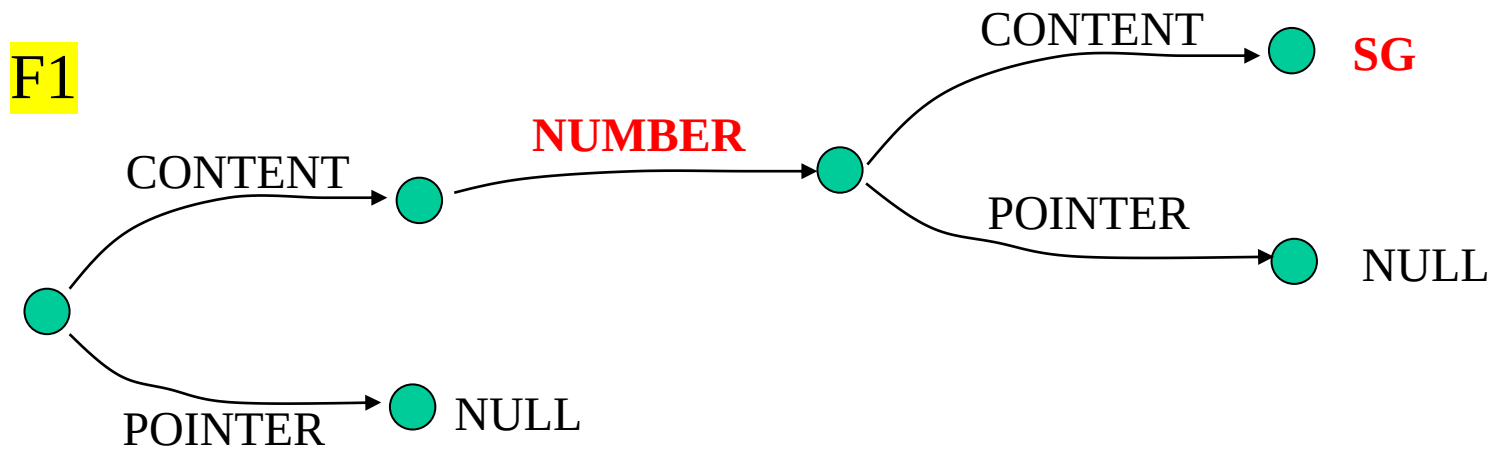
(2) f1，f2是原子值，二者相等

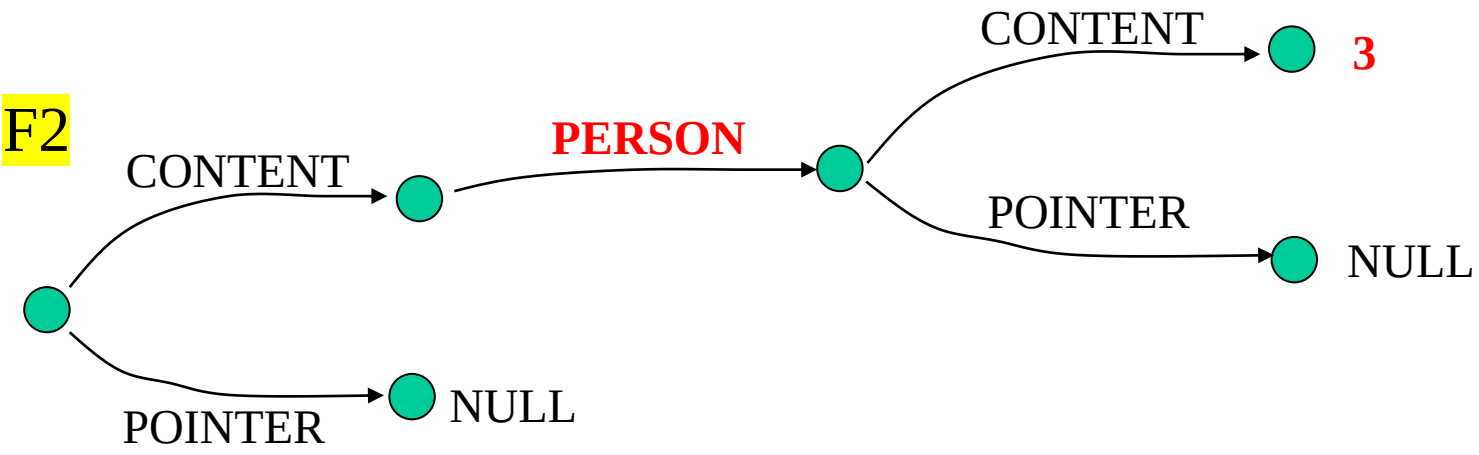
(3) f1和f2均为复杂特征

f1和f2冲突，合一失败

例1

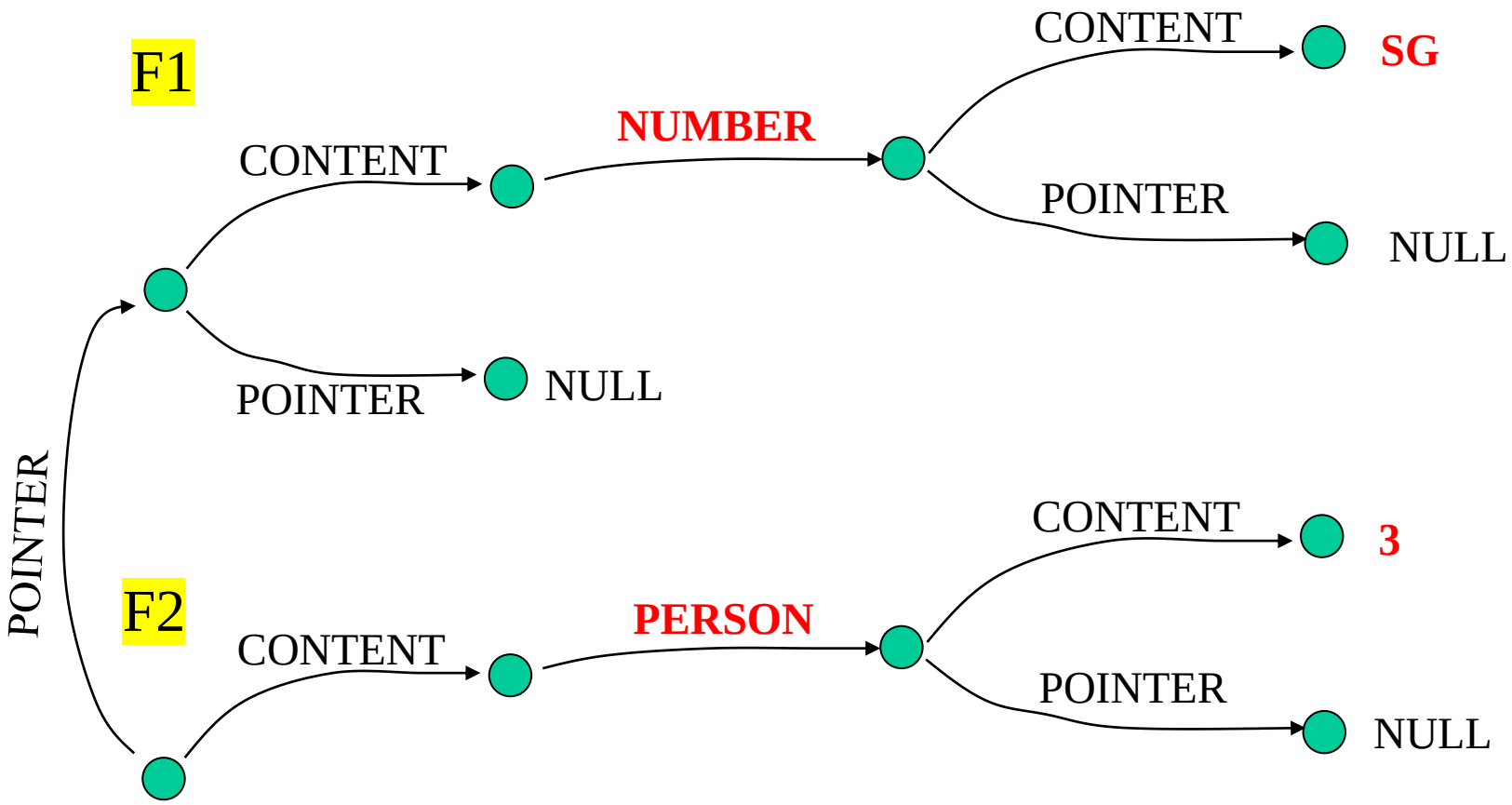
$$\begin{pmatrix} \text{NUMBER} & \text{SG} \end{pmatrix} \cup \begin{pmatrix} \text{PERSON} & 3 \end{pmatrix} = \begin{pmatrix} \text{NUMBER} & \text{SG} \\ \text{PERSON} & 3 \end{pmatrix}$$





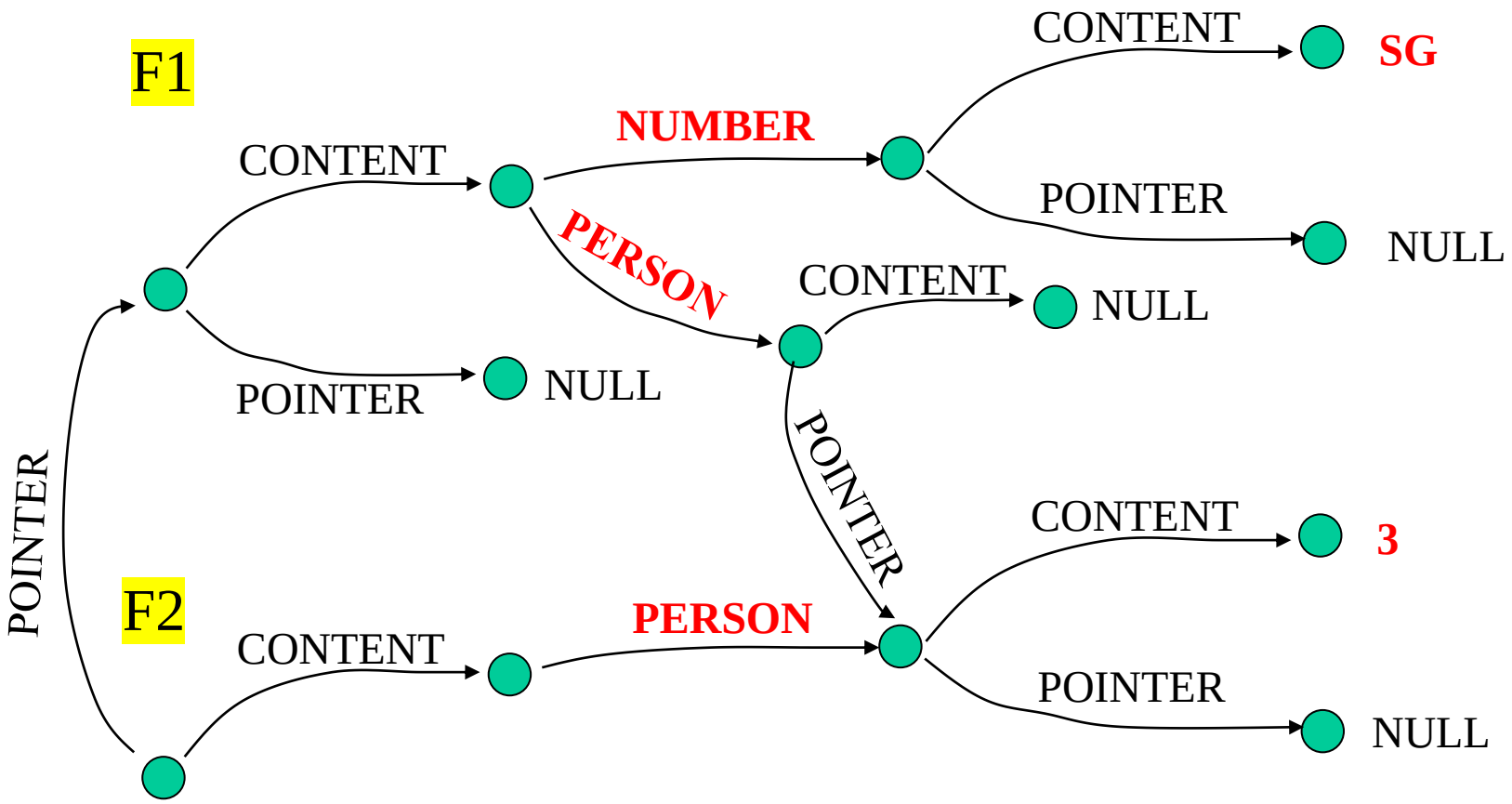
例1

$$\begin{bmatrix} \text{NUMBER} & \text{SG} \end{bmatrix} \cup \begin{bmatrix} \text{PERSON} & 3 \end{bmatrix} = \begin{pmatrix} \text{NUMBER} & \text{SG} \\ \text{PERSON} & 3 \end{pmatrix}$$



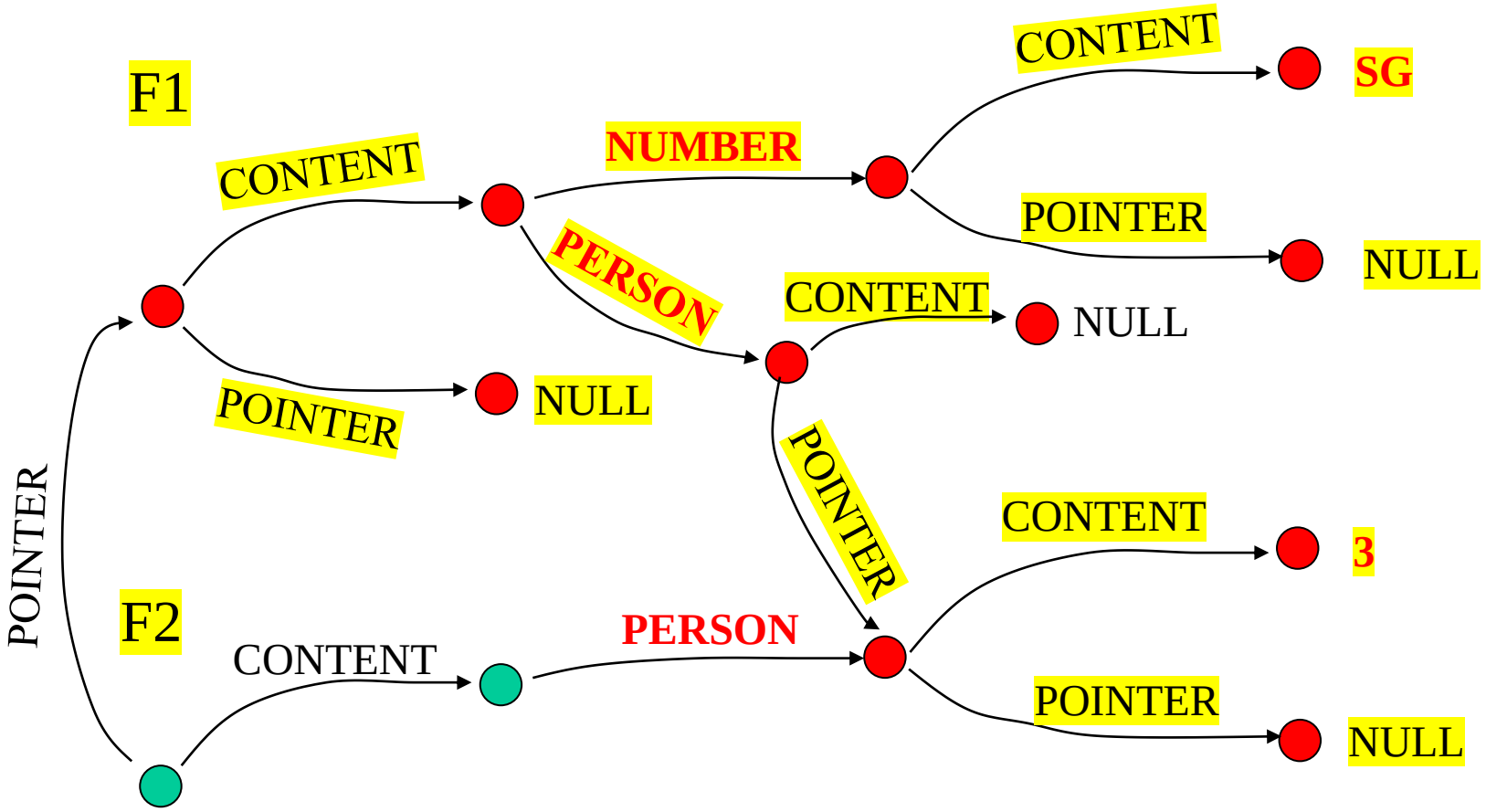
例1

$$\begin{bmatrix} \text{NUMBER} & \text{SG} \end{bmatrix} \cup \begin{bmatrix} \text{PERSON} & 3 \end{bmatrix} = \begin{bmatrix} \text{NUMBER} & \text{SG} \\ \text{PERSON} & 3 \end{bmatrix}$$

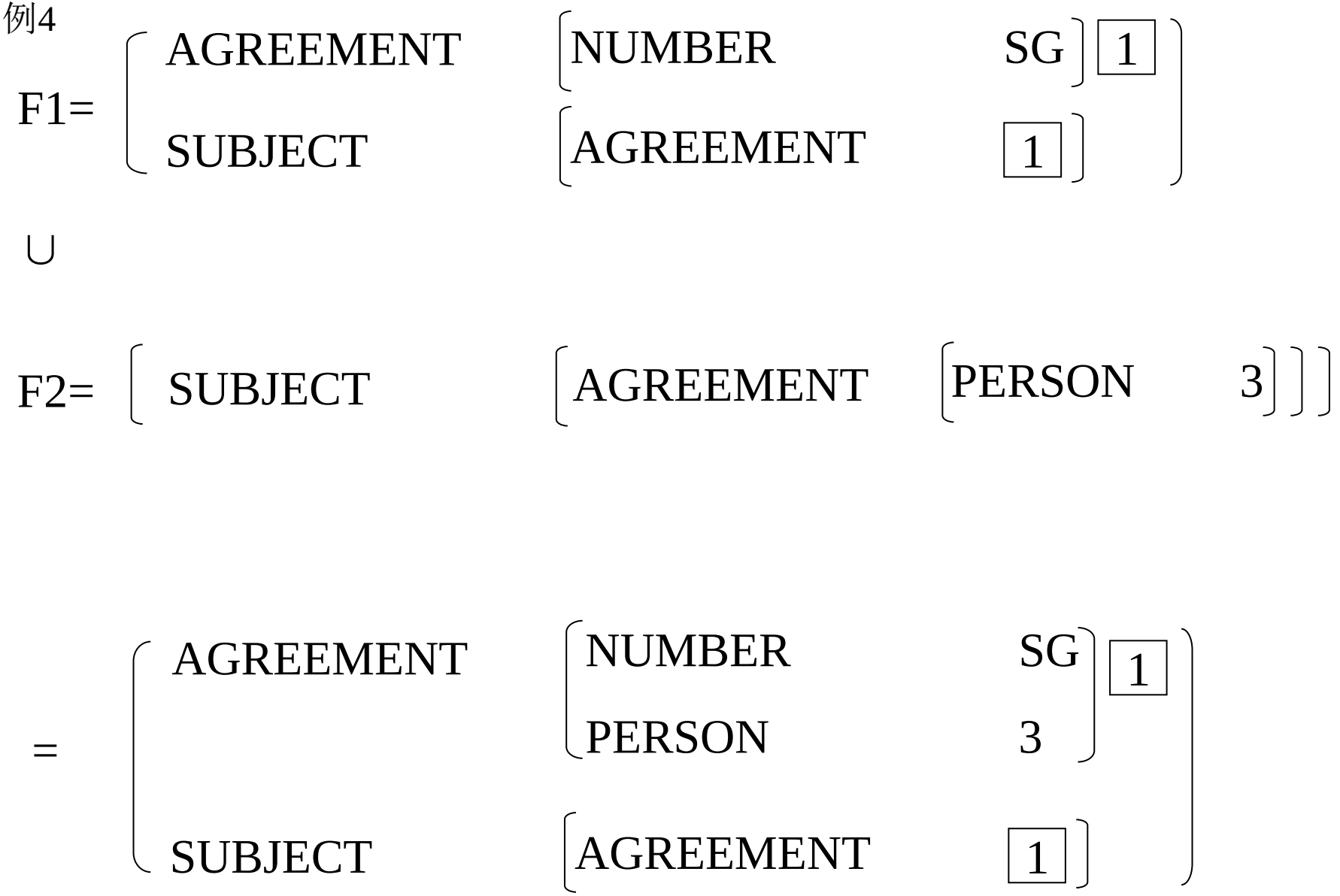


例1

$$\begin{bmatrix} \text{NUMBER} & \text{SG} \end{bmatrix} \cup \begin{bmatrix} \text{PERSON} & 3 \end{bmatrix} = \begin{pmatrix} \text{NUMBER} & \text{SG} \\ \text{PERSON} & 3 \end{pmatrix}$$

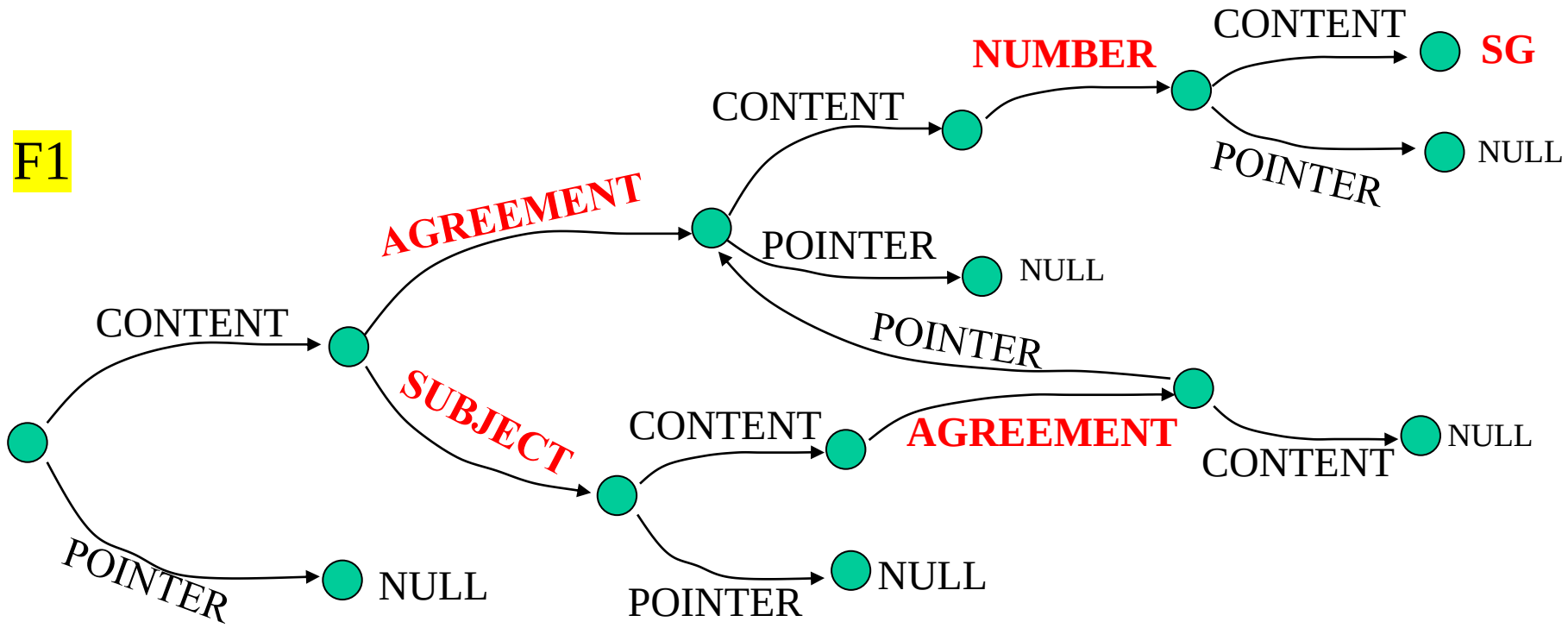


例4


F1=
AGREEMENT
NUMBER
SG
1
SUBJECT
AGREEMENT
1
∪
F2=
SUBJECT
AGREEMENT
PERSON
3
=
AGREEMENT
NUMBER
SG
PERSON
3
1
SUBJECT
AGREEMENT
1

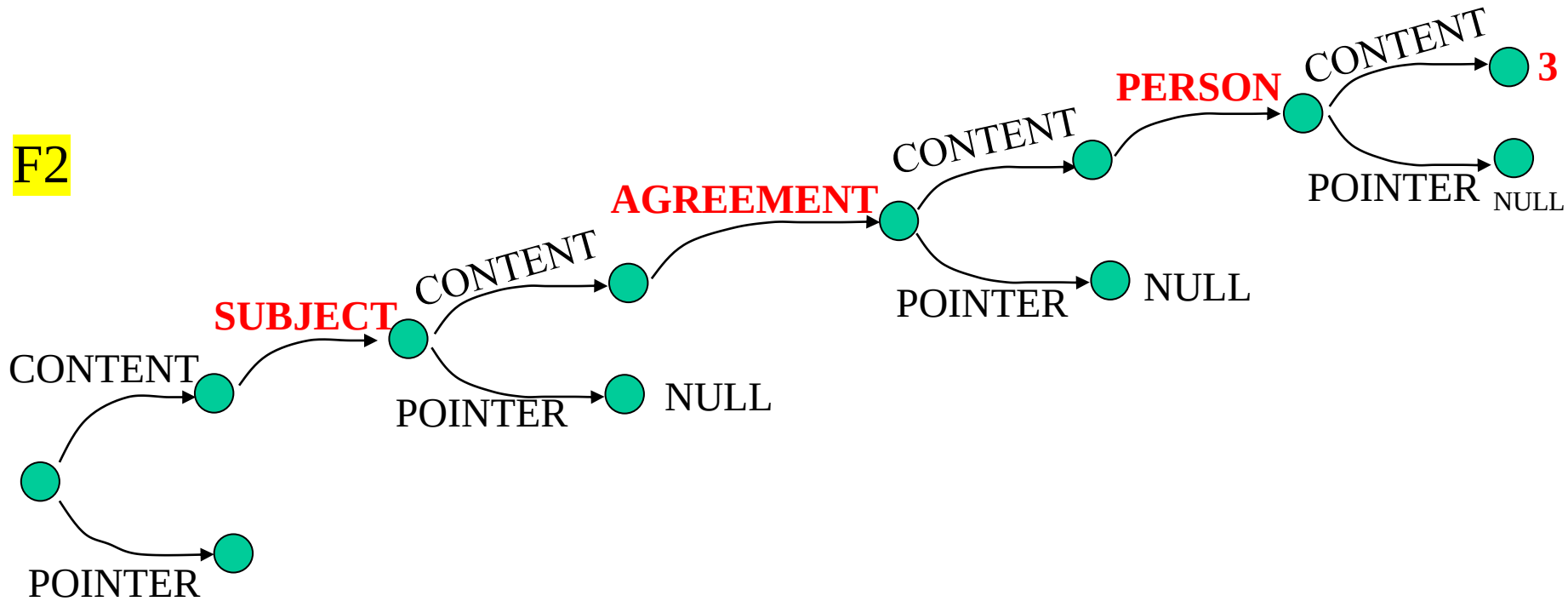
例4

$$F1=\begin{bmatrix} AGREEMENT & \begin{bmatrix} NUMBER & SG & \boxed{1} \end{bmatrix} \\ SUBJECT & \begin{bmatrix} AGREEMENT & \boxed{1} \end{bmatrix} \end{bmatrix}$$

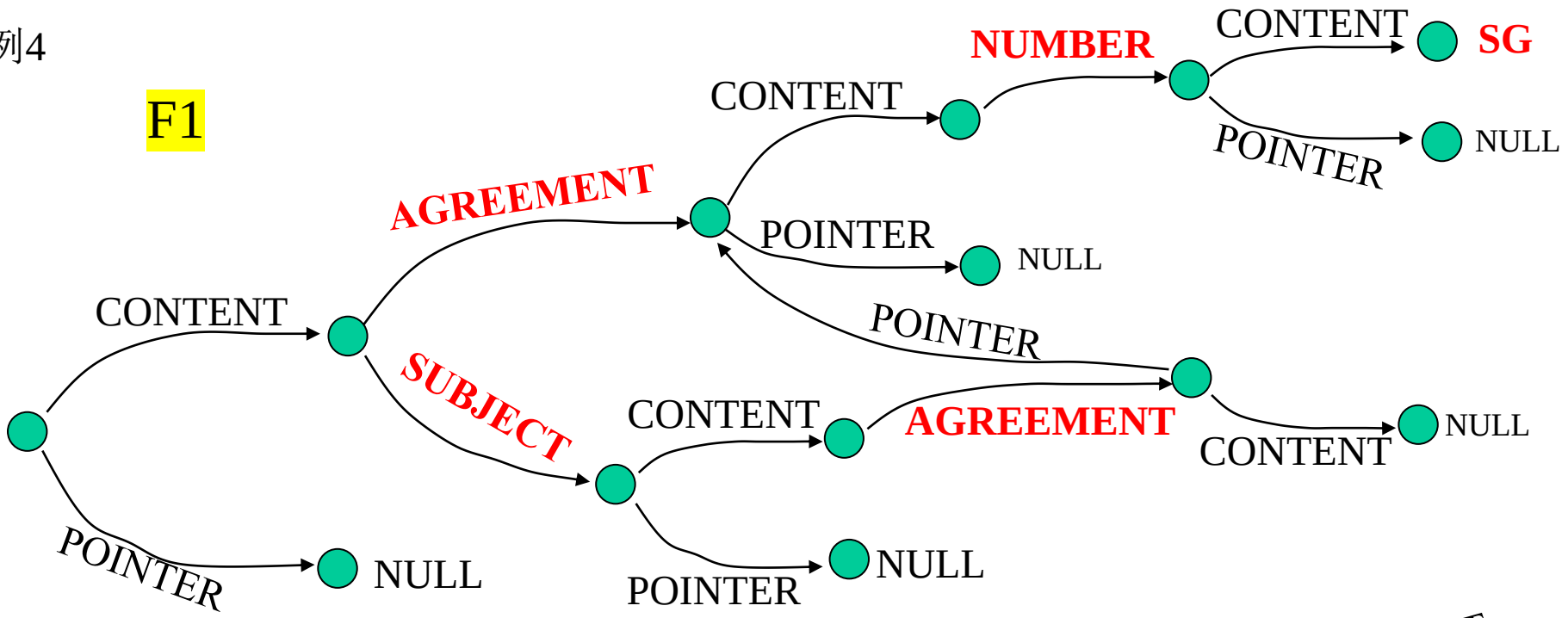


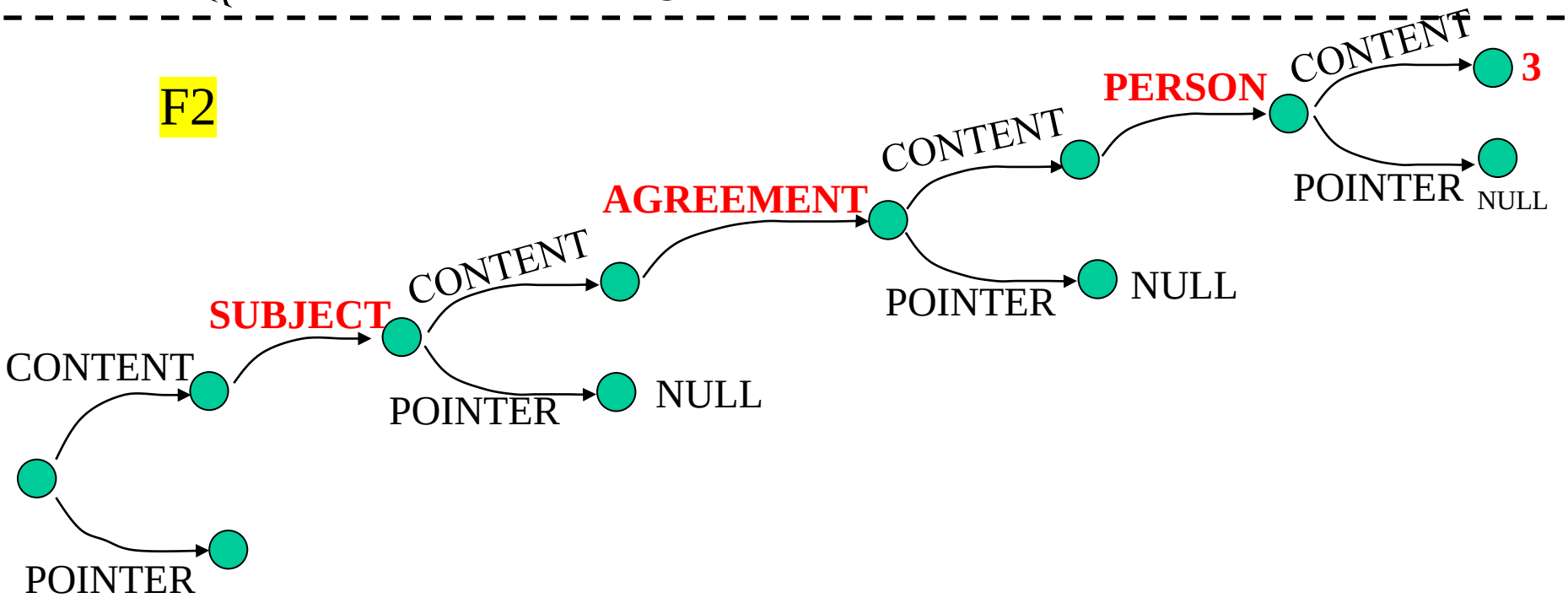
例4

F2= [SUBJECT [AGREEMENT [PERSON 3]]]

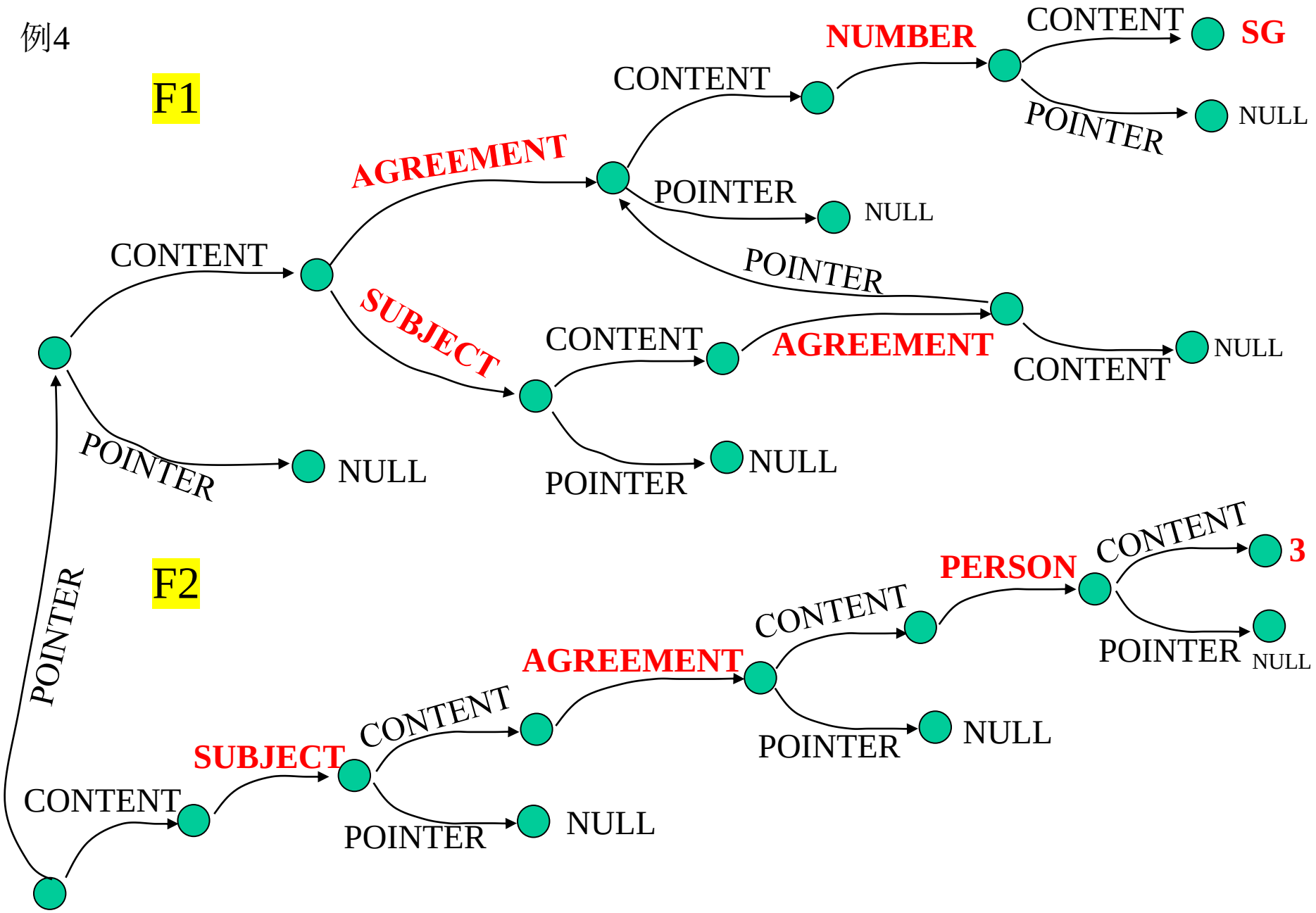


例4

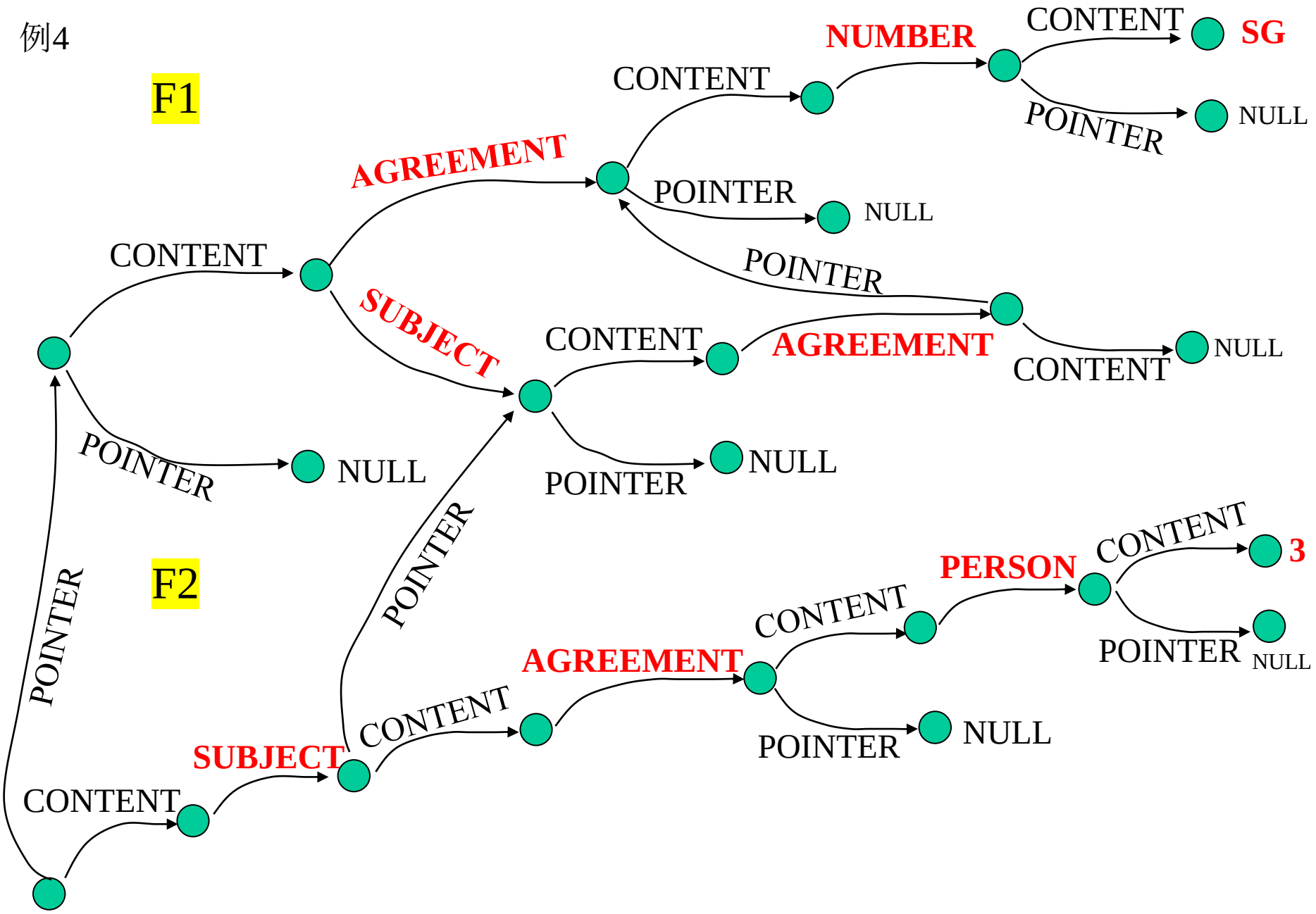

F1
NUMBER
CONTENT
SG
CONTENT
NULL
POINTER
AGREEMENT
POINTER
NULL
CONTENT
POINTER
SUBJECT
CONTENT
AGREEMENT
NULL
CONTENT
POINTER
NULL
POINTER
NULL


F2
PERSON
CONTENT
3
CONTENT
POINTER
NULL
AGREEMENT
CONTENT
POINTER
NULL
SUBJECT
CONTENT
CONTENT
POINTER
NULL
POINTER

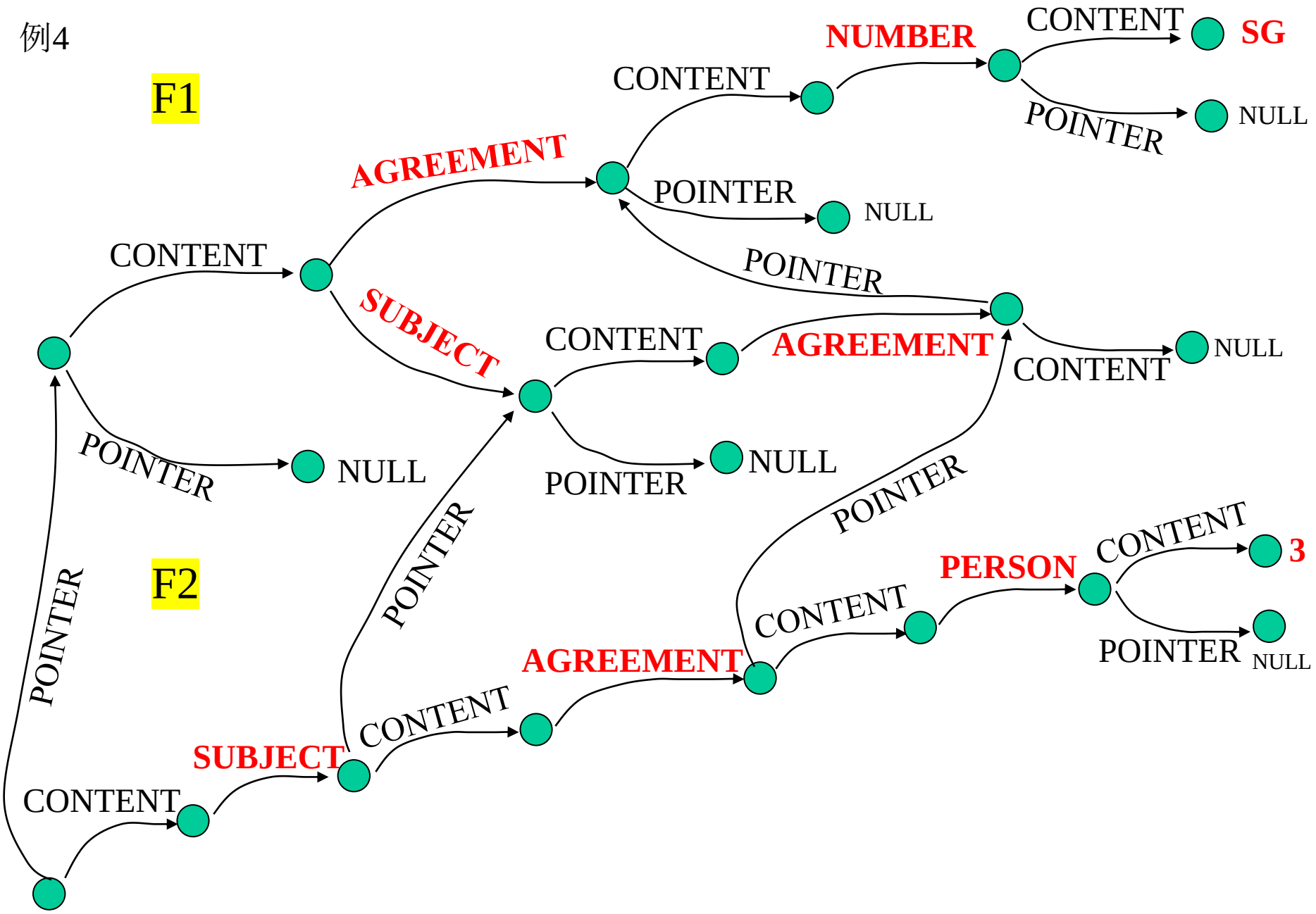
例4


F1
NUMBER
CONTENT
SG
CONTENT
POINTER
NULL
AGREEMENT
POINTER
NULL
CONTENT
POINTER
SUBJECT
CONTENT
AGREEMENT
CONTENT
NULL
POINTER
NULL
POINTER
NULL
F2
PERSON
CONTENT
3
CONTENT
POINTER
NULL
AGREEMENT
CONTENT
POINTER
NULL
SUBJECT
CONTENT
POINTER
NULL
CONTENT
POINTER

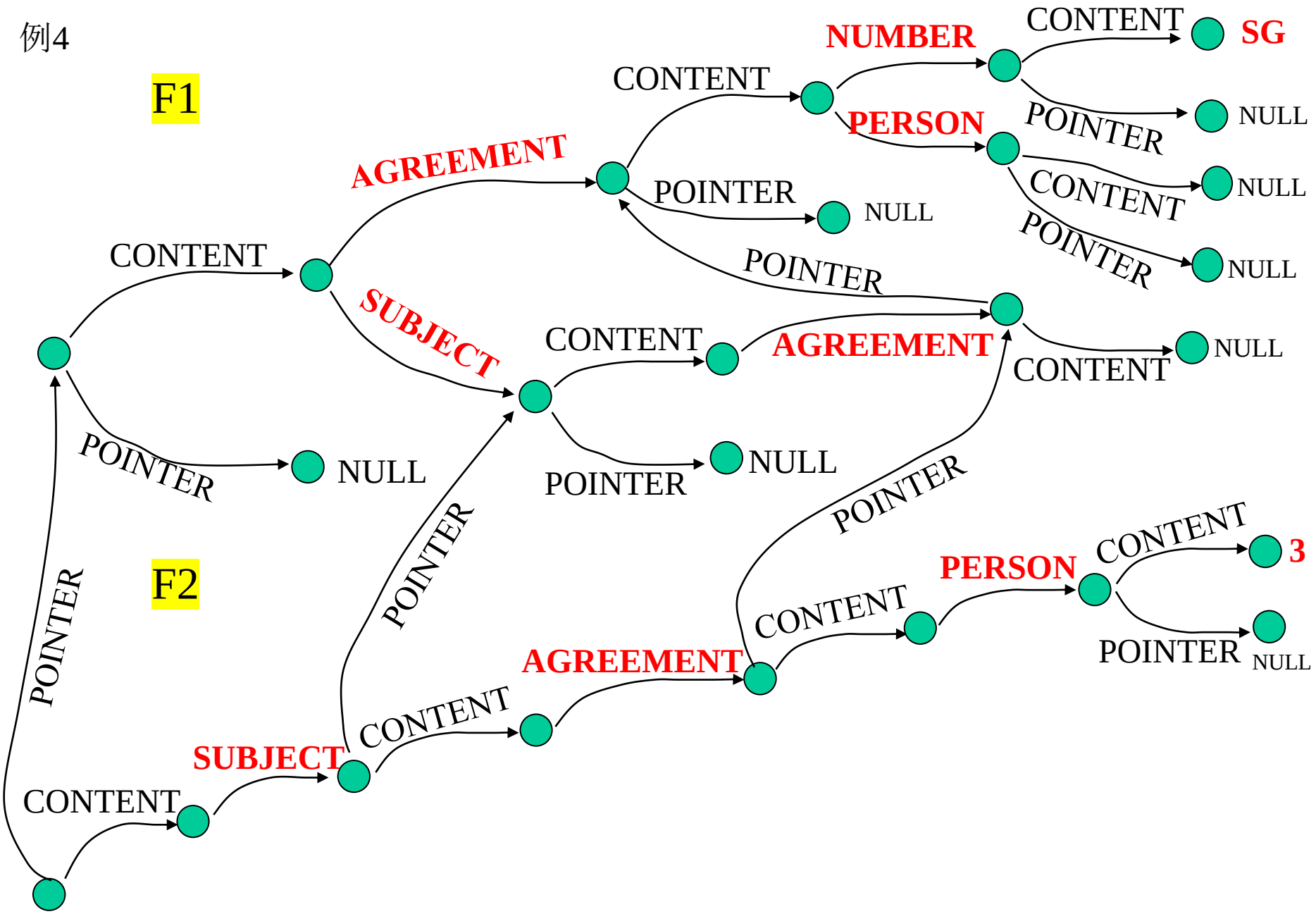
例4


F1
F2
CONTENT
NUMBER
SG
NULL
POINTER
AGREEMENT
SUBJECT
PERSON
3

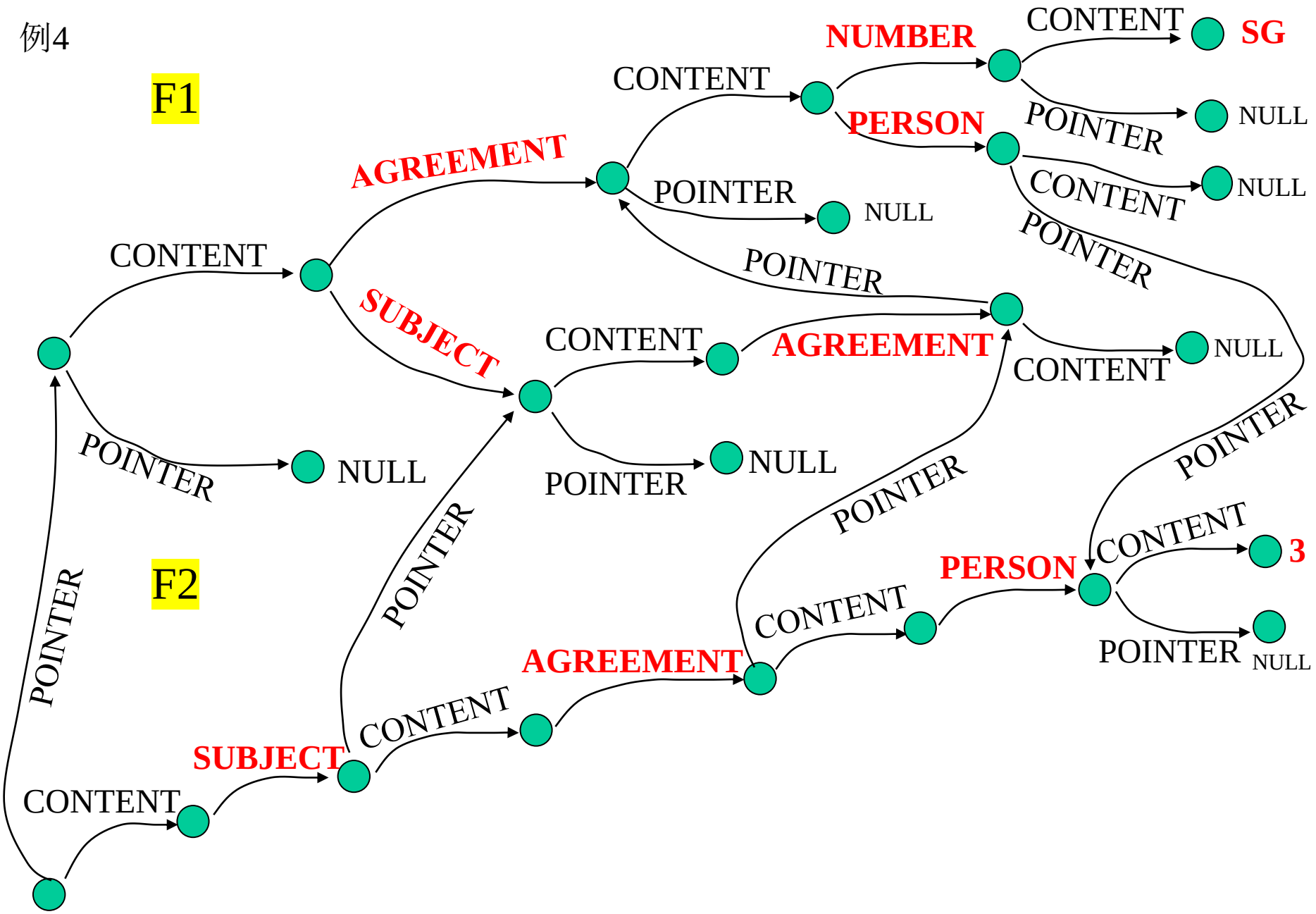
例4


F1
F2
NUMBER
CONTENT
SG
CONTENT
NULL
POINTER
AGREEMENT
POINTER
NULL
CONTENT
POINTER
SUBJECT
CONTENT
AGREEMENT
NULL
CONTENT
POINTER
NULL
POINTER
NULL
POINTER
POINTER
PERSON
CONTENT
3
CONTENT
POINTER
NULL
AGREEMENT
POINTER
SUBJECT
CONTENT
CONTENT

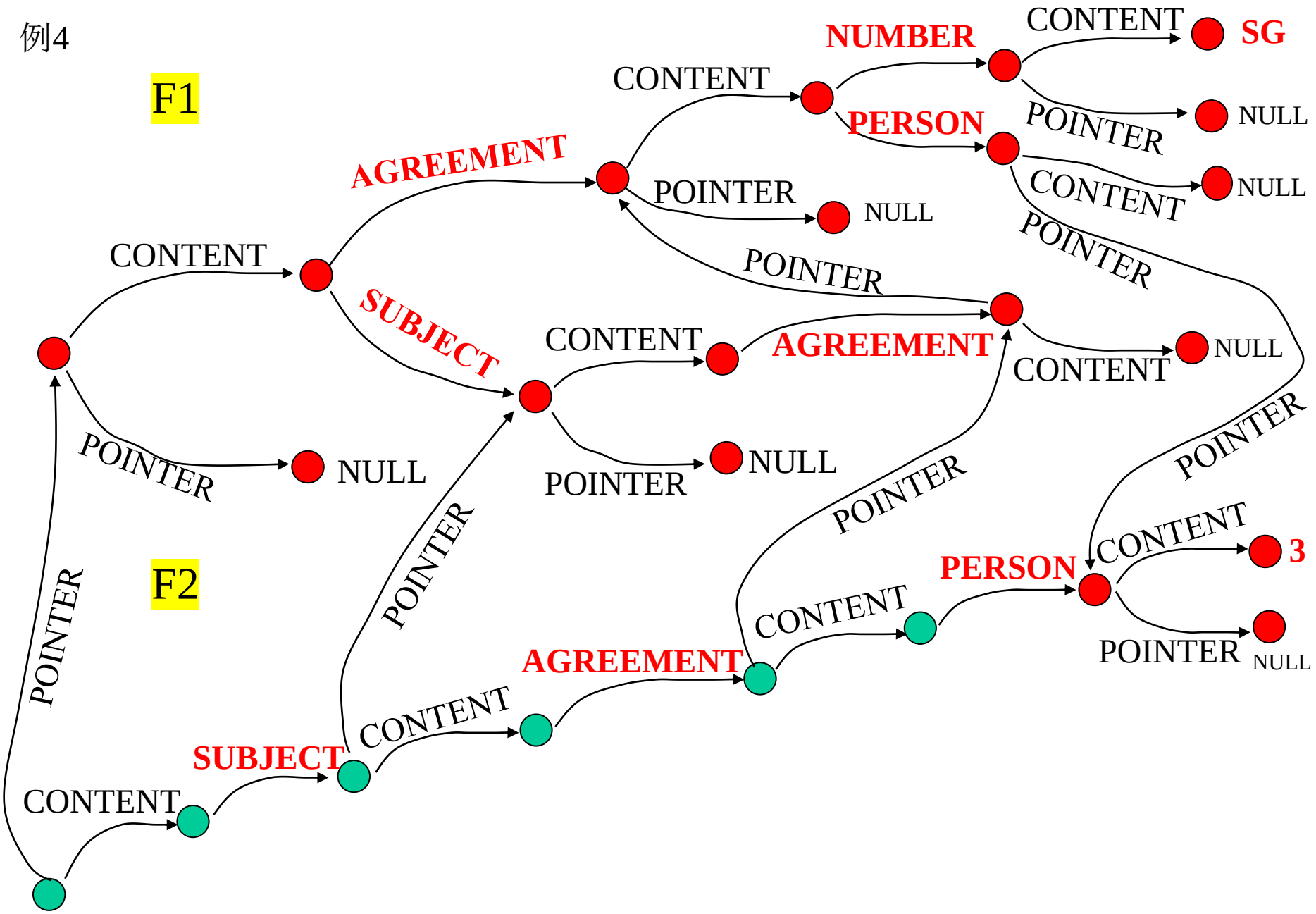
例4


F1
F2
CONTENT
AGREEMENT
SUBJECT
POINTER
NULL
NUMBER
PERSON
SG
3

例4


F1
F2
NUMBER
CONTENT
SG
CONTENT
NULL
PERSON
POINTER
AGREEMENT
POINTER
NULL
CONTENT
NULL
CONTENT
POINTER
POINTER
CONTENT
SUBJECT
CONTENT
AGREEMENT
NULL
CONTENT
POINTER
NULL
POINTER
NULL
POINTER
POINTER
POINTER
CONTENT
3
PERSON
CONTENT
POINTER
NULL
AGREEMENT
CONTENT
SUBJECT
CONTENT
POINTER
CONTENT

例4


F1
F2
CONTENT
NUMBER
SG
PERSON
NULL
AGREEMENT
POINTER
SUBJECT
3