

INF5180

**Conception et exploitation d'une
base de données**

Cours 3

Zied Zaier, PhD

Département d'informatique
Université du Québec à Montréal

Cours 3

THÉORIE DE LA NORMALISATION

**CE DOCUMENT EST INSPIRÉ DES TRAVAUX DES PROFESSEURS ROBERT GODIN,
VALTCHEV PETKO ET FATIHA SADAT.**

Sommaire

- I. Rappel
- II. Décomposition sans perte
- III. Dépendance fonctionnelle
 - Interférence et choix des dépendances fonctionnelles
 - Axiomes d'Armstrong
 - Représentation des dépendances fonctionnelles par un diagramme à bulles
 - Couverture minimale d'un ensemble de dépendances fonctionnelles
 - Dépendances fonctionnelles et clés candidates
- IV. Normalisation

Exemple de mauvaise conception

Redondance de données

Table *Vente*

noCommande	dateCommande	noClient	nomClient	noTéléphon	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Luc Sansom	(999)999-9999	10	Cèdre en boule	10.99	10
1	01/06/2000	10	Luc Sansom	(999)999-9999	70	Herbe à puce	10.99	5
1	01/06/2000	10	Luc Sansom	(999)999-9999	90	Pommier	25.99	1
2	02/06/2000	20	D o l l a r d Tremblay	(888)888-8888	40	Epinette bleue	25.99	2
2	02/06/2000	20	D o l l a r d Tremblay	(888)888-8888	95	Génévrier	15.99	3
3	02/06/2000	10	Luc Sansom	(999)999-9999	20	Sapin	12.99	1
4	05/06/2000	10	Luc Sansom	(999)999-9999	40	Epinette bleue	25.99	1
4	05/06/2000	10	Luc Sansom	(999)999-9999	50	Chêne	22.99	1
5	09/07/2000	30	Lin Bô	(777)777-7777	70	Herbe à puce	10.99	3
5	09/07/2000	30	Lin Bô	(777)777-7777	10	Cèdre en boule	10.99	5
5	09/07/2000	30	Lin Bô	(777)777-7777	20	Sapin	12.99	5
6	09/07/2000	20	D o l l a r d Tremblay	(888)888-8888	10	Cèdre en boule	10.99	5
6	09/07/2000	20	D o l l a r d Tremblay	(888)888-8888	40	Epinette bleue	25.99	1
7	15/07/2000	40	Jean Leconte	(666)666-6666	50	Chêne	22.99	1
7	15/07/2000	40	Jean Leconte	(666)666-6666	95	Génévrier	15.99	2
8	15/07/2000	40	Jean Leconte	(666)666-6666	20	Sapin	12.99	3

- *Vente (noCl, nomCl, noTél, noCom, dateCom, noArt, desc, prixU, quantité)*

Solution : décomposition

Table <i>Vente</i>								
noCommande	dateCommande	noClient	nomClient	noTéléphone	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Luc Sansom	(999)999-9999	10	Cèdre en boule	10.99	10
1	01/06/2000	10	Luc Sansom	(999)999-9999	70	Herbe à puce	10.99	5
1	01/06/2000	10	Luc Sansom	(999)999-9999	90	Pommier	25.99	1
...								

Table <i>Client</i>		
noClient	nomClient	noTéléphone
10	Luc Sansom	(999)999-9999
20	Dollard Tremblay	(888)888-8888
30	Lin Bô	(777)777-7777
40	Jean Leconte	(666)666-6666
50	Hafedh Lajoie	(555)555-5555
60	Comtesse Hasek	(666)666-6666
70	Coco McPoulet	(444)444-4419
80	Dollard Tremblay	(333)333-3333

Table <i>VenteReste</i>							
noCommande	dateCommande	noClient	noArticle	description	prixUnitaire	quantité	
1	01/06/2000	10	10	Cèdre en boule	10.99	10	
1	01/06/2000	10	70	Herbe à puce	10.99	5	
1	01/06/2000	10	90	Pommier	25.99	1	
2	02/06/2000	20	40	Epinette bleue	25.99	2	
2	02/06/2000	20	95	Génévrier	15.99	3	
3	02/06/2000	10	20	Sapin	12.99	1	
4	05/06/2000	10	40	Epinette bleue	25.99	1	
4	05/06/2000	10	50	Chêne	22.99	1	
5	09/07/2000	30	70	Herbe à puce	10.99	3	
5	09/07/2000	30	10	Cèdre en boule	10.99	5	
5	09/07/2000	30	20	Sapin	12.99	5	
6	09/07/2000	20	10	Cèdre en boule	10.99	5	
6	09/07/2000	20	40	Epinette bleue	25.99	1	
7	15/07/2000	40	50	Chêne	22.99	1	
7	15/07/2000	40	95	Génévrier	15.99	2	
8	15/07/2000	40	20	Sapin	12.99	3	

- Client (*noCl*, *nomCl*, *noTél*)
- *VenteReste*(*noCom*, *dateCom*, *noCl*, *noArt*, *desc*, *prixU*, *quantité*)

Décomposition sans perte

- La décomposition ne doit pas s'effectuée au hasard
 - Perte d'information
 - Impossible de reconstruire la table de départ
- Décompositions acceptables:
 - Décomposition binaire
 - Sans perte

Décomposition binaire

- Une décomposition binaire d'une table $T(A_1, A_2, \dots, A_n)$ en deux tables $T_1(B_1, B_2, \dots, B_m)$ et $T_2(C_1, C_2, \dots, C_p)$ si:

$$\{A_1, A_2, \dots, A_n\} = \{B_1, B_2, \dots, B_m\} \cup \{C_1, C_2, \dots, C_p\}$$

$$T_1 = \pi B_1, B_2, \dots, B_m (T)$$

$$T_2 = \pi C_1, C_2, \dots, C_p (T)$$

Décomposition binaire

Table Client		
noClient	nomClient	noTéléphone
10	Luc Sansom	(999)999-9999
20	Dollard Tremblay	(888)888-8888
30	Lin Bô	(777)777-7777
40	Jean Leconte	(666)666-6666
50	Hafedh Lajoie	(555)555-5555
60	Comtesse Hasek	(666)666-6666
70	Coco McPoulet	(444)444-4419
80	Dollard Tremblay	(333)333-3333

Table VenteAvecPerte

noCommande	dateCommande	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Cèdre en boule	10.99	10
1	01/06/2000	70	Herbe à puce	10.99	5
1	01/06/2000	90	Pommier	25.99	1
2	02/06/2000	40	Epinette bleue	25.99	2
2	02/06/2000	95	Génévrier	15.99	3
3	02/06/2000	20	Sapin	12.99	1
4	05/06/2000	40	Epinette bleue	25.99	1
4	05/06/2000	50	Chêne	22.99	1
5	09/07/2000	70	Herbe à puce	10.99	3
5	09/07/2000	10	Cèdre en boule	10.99	5
5	09/07/2000	20	Sapin	12.99	5
6	09/07/2000	10	Cèdre en boule	10.99	5
6	09/07/2000	40	Epinette bleue	25.99	1
7	15/07/2000	50	Chêne	22.99	1
7	15/07/2000	95	Génévrier	15.99	2
8	15/07/2000	20	Sapin	12.99	3

VenteAvecPerte(noCom, dateCom, noArt, desc, prixU, quantité)

Client (noCl, nomCl, noTél)

Décomposition sans perte

- Une décomposition binaire d'une table T en deux tables T_1 et T_2 est sans perte si :
 - $T = T_1 \mid T_2$
- Exemple :
 - $Vente = Client \mid VenteReste$

| est le signe de jointure naturelle \otimes

Exemple de décomposition avec perte

- *Vente ≠ Client | VenteAvecPerte*

Table <i>Client</i>		
noClient	nomClient	noTéléphone
10	Luc Sansom	(999)999-9999
20	Dollard Tremblay	(888)888-8888
30	Lin Bô	(777)777-7777
40	Jean Leconte	(666)666-6666
50	Hafedh Lajoie	(555)555-5555
60	Comtesse Hasek	(666)666-6666
70	Coco McPoulet	(444)444-4419
80	Dollard Tremblay	(333)333-3333

Table <i>VenteAvecPerte</i>					
noCommande	dateCommande	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Cèdre en boule	10.99	10
1	01/06/2000	70	Herbe à puce	10.99	5
1	01/06/2000	90	Pommier	25.99	1
2	02/06/2000	40	Epinette bleue	25.99	2
2	02/06/2000	95	Génévrier	15.99	3
3	02/06/2000	20	Sapin	12.99	1
4	05/06/2000	40	Epinette bleue	25.99	1
4	05/06/2000	50	Chêne	22.99	1
5	09/07/2000	70	Herbe à puce	10.99	3
5	09/07/2000	10	Cèdre en boule	10.99	5
5	09/07/2000	20	Sapin	12.99	5
6	09/07/2000	10	Cèdre en boule	10.99	5
6	09/07/2000	40	Epinette bleue	25.99	1
7	15/07/2000	50	Chêne	22.99	1
7	15/07/2000	95	Génévrier	15.99	2
8	15/07/2000	20	Sapin	12.99	3

Solution : décomposition

Table <i>Vente</i>								
noCommande	dateCommande	noClient	nomClient	noTéléphone	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Luc Sansom	(999)999-9999	10	Cèdre en boule	10.99	10
1	01/06/2000	10	Luc Sansom	(999)999-9999	70	Herbe à puce	10.99	5
1	01/06/2000	10	Luc Sansom	(999)999-9999	90	Pommier	25.99	1
...								

Table <i>Client</i>		
noClient	nomClient	noTéléphone
10	Luc Sansom	(999)999-9999
20	Dollard Tremblay	(888)888-8888
30	Lin Bô	(777)777-7777
40	Jean Leconte	(666)666-6666
50	Hafedh Lajoie	(555)555-5555
60	Comtesse Hasek	(666)666-6666
70	Coco McPoulet	(444)444-4419
80	Dollard Tremblay	(333)333-3333

Table <i>VenteReste</i>							
noCommande	dateCommande	noClient	noArticle	description	prixUnitaire	quantité	
1	01/06/2000	10	10	Cèdre en boule	10.99	10	
1	01/06/2000	10	70	Herbe à puce	10.99	5	
1	01/06/2000	10	90	Pommier	25.99	1	
2	02/06/2000	20	40	Epinette bleue	25.99	2	
2	02/06/2000	20	95	Génévrier	15.99	3	
3	02/06/2000	10	20	Sapin	12.99	1	
4	05/06/2000	10	40	Epinette bleue	25.99	1	
4	05/06/2000	10	50	Chêne	22.99	1	
5	09/07/2000	30	70	Herbe à puce	10.99	3	
5	09/07/2000	30	10	Cèdre en boule	10.99	5	
5	09/07/2000	30	20	Sapin	12.99	5	
6	09/07/2000	20	10	Cèdre en boule	10.99	5	
6	09/07/2000	20	40	Epinette bleue	25.99	1	
7	15/07/2000	40	50	Chêne	22.99	1	
7	15/07/2000	40	95	Génévrier	15.99	2	
8	15/07/2000	40	20	Sapin	12.99	3	

- Client (*noCl*, *nomCl*, *noTél*)
- *VenteReste*(*noCom*, *dateCom*, *noCl*, *noArt*, *desc*, *prixU*, *quantité*)

Redondance des données

- Critère de décomposition sans perte pour **Vente(noClient, nomClient, noTéléphone, noCommande, dateCommande, noArticle, description, prixUnitaire, quantité)**
 - **T(noClient, nomClient, noTéléphone)**
 - **S(noClient, noCommande, dateCommande, noArticle, description, prixUnitaire, quantité)**
- Non suffisant
 - Théorie de la normalisation
 - résoudre le problème de redondance
 - notion de dépendance fonctionnelle

Dépendance fonctionnelle

- Notée $A_1, A_2, \dots, A_n \rightarrow^T B_1, B_2, \dots, B_m$
 - pour toutes les instances de la table T, lorsque deux lignes ont les mêmes valeurs pour les colonnes A1, A2, ..., An \Rightarrow elles ont nécessairement les mêmes valeurs pour les colonnes B1, B2, ..., Bm
- Déterminant : la partie gauche d'une dépendance (A_1, A_2, \dots, A_n)
- Vocabulaire:
 - A1, A2, ..., An détermine B1, B2, ..., Bm
 - B1, B2, ..., Bm dépend de A1, A2, ..., An

Exemples pour Vente

Table *Vente*

noCommande	dateCommande	noClient	nomClient	noTéléphone	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Luc Sansom	(999)999-9999	10	Cèdre en boule	10.99	10
1	01/06/2000	10	Luc Sansom	(999)999-9999	70	Herbe à puce	10.99	5
1	01/06/2000	10	Luc Sansom	(999)999-9999	90	Pommier	25.99	1
2	02/06/2000	20	Dollard Tremblay	(888)888-8888	40	Epinette bleue	25.99	2
2	02/06/2000	20	Dollard Tremblay	(888)888-8888	95	Génévrier	15.99	3
3	02/06/2000	10	Luc Sansom	(999)999-9999	20	Sapin	12.99	1
4	05/06/2000	10	Luc Sansom	(999)999-9999	40	Epinette bleue	25.99	1
4	05/06/2000	10	Luc Sansom	(999)999-9999	50	Chêne	22.99	1
5	09/07/2000	30	Lin Bô	(777)777-7777	70	Herbe à puce	10.99	3
5	09/07/2000	30	Lin Bô	(777)777-7777	10	Cèdre en boule	10.99	5
5	09/07/2000	30	Lin Bô	(777)777-7777	20	Sapin	12.99	5
6	09/07/2000	20	Dollard Tremblay	(888)888-8888	10	Cèdre en boule	10.99	5
6	09/07/2000	20	Dollard Tremblay	(888)888-8888	40	Epinette bleue	25.99	1
7	15/07/2000	40	Jean Leconte	(666)666-6666	50	Chêne	22.99	1
7	15/07/2000	40	Jean Leconte	(666)666-6666	95	Génévrier	15.99	2
8	15/07/2000	40	Jean Leconte	(666)666-6666	20	Sapin	12.99	3

Les dépendances fonctionnelles suivantes sont valides pour la table Vente:

noClient → nomClient, noTéléphone

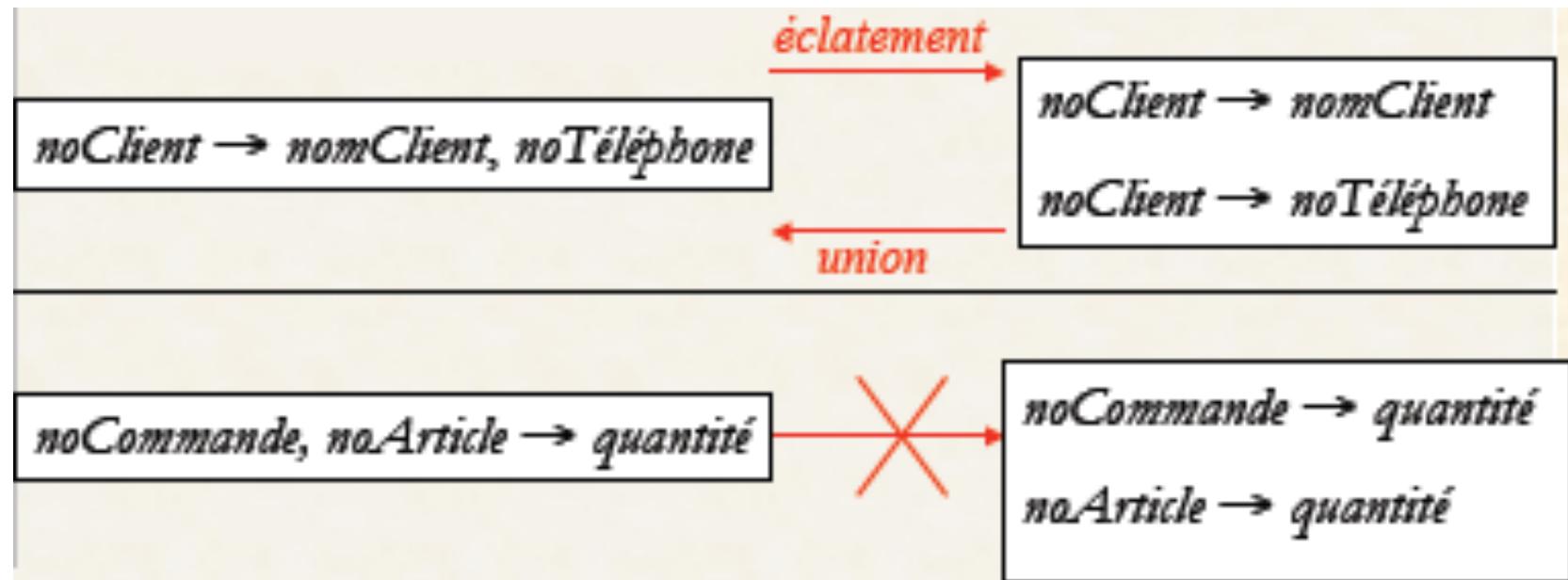
noCommande → noClient, dateCommande

noCommande, noArticle → quantité

noArticle → description, prixUnitaire

Inférence et choix des dépendances fonctionnelles

- Considérer toutes les dépendances ?
- Sinon lesquelles ?
- Équivalences logiques



Dépendance élémentaire

- Dépendance fonctionnelle **triviale**
 - Partie droite incluse dans la partie gauche (le déterminant)

noCommande, noArticle, description → noCommande, noArticle

- Dépendance fonctionnelle **complètement non triviale**
 - Aucune colonne de la partie droite n'apparaît dans la partie gauche

noClient → noClient, nomClient

Dépendance élémentaire

- Dépendance fonctionnelle **pleine** (« *full* »)

– Partie gauche minimale

~~$noClient, noCommande \rightarrow nomClient$~~

- Dépendance fonctionnelle $X \rightarrow a$ élémentaire:

- complètement non triviale, pleine et $|a| = 1$ (a est un seul attribut)

$noClient \rightarrow nomClient$

Dépendances superflues (redondantes)

$noClient \rightarrow nomClient, noTéléphone$

$noCommande \rightarrow noClient, dateCommande$

~~$noCommande \rightarrow nomClient$~~

Par transitivité

- $noCommande \rightarrow noClient$ est déductible de la deuxième dépendance par la règle d'éclatement et $noClient \rightarrow nomClient$ est déductible de la première dépendance par la règle d'éclatement.
 $noCommande \rightarrow noClient$ correspond à l'axiome de transitivité.

Axiomes d'Armstrong

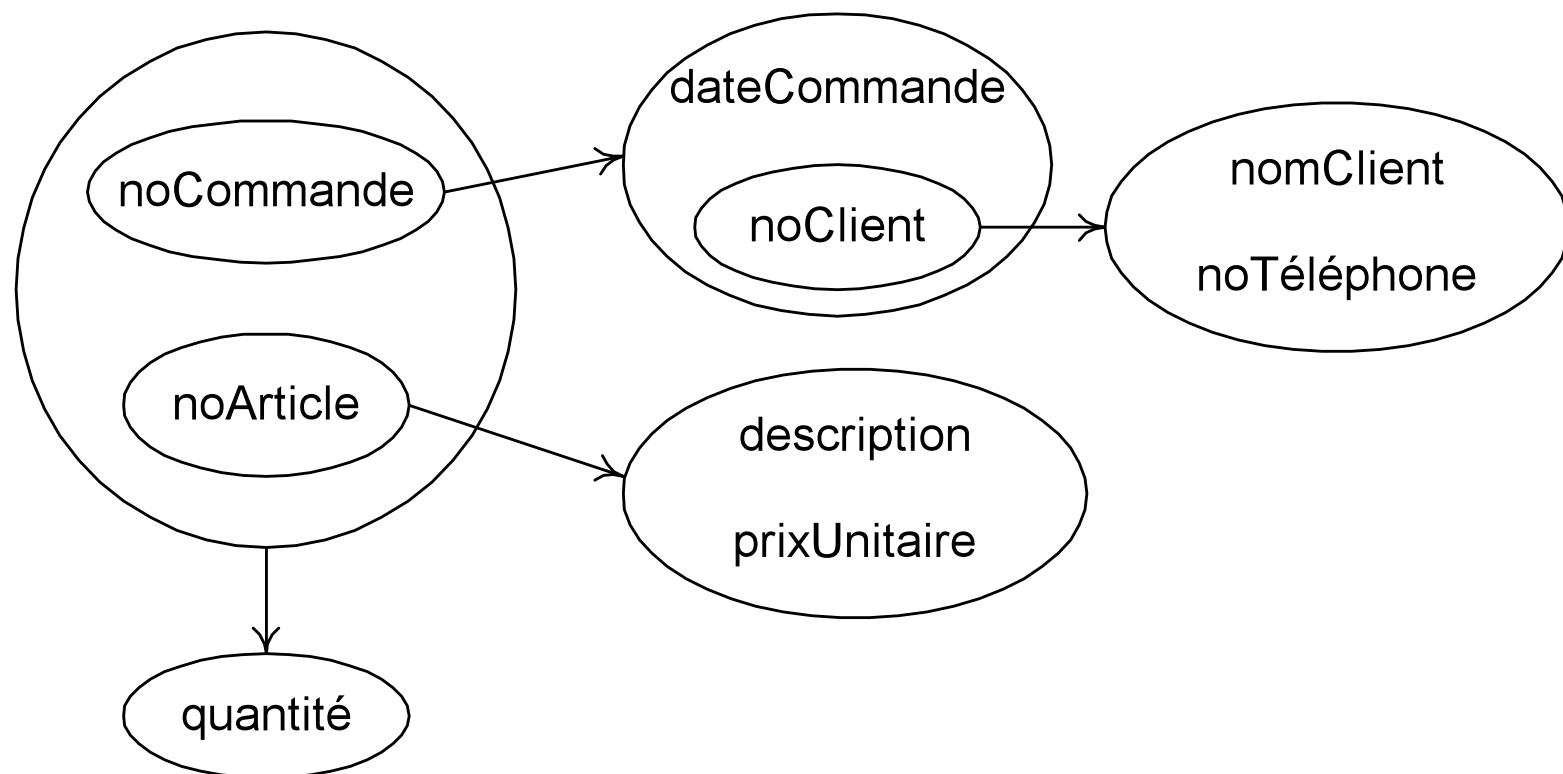
- ***A₁. Réflexivité***
 - Si $\{B_1, B_2, \dots, B_m\} \subseteq \{A_1, A_2, \dots, A_n\}$ alors
 - $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$
- ***A₂. Augmentation***
 - Si $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ et
 - $\{D_1, D_2, \dots, D_p\} \subseteq \{C_1, C_2, \dots, C_r\}$ alors
 - $A_1, A_2, \dots, A_n, C_1, C_2, \dots, C_r \rightarrow B_1, B_2, \dots, B_m, D_1, D_2, \dots, D_p$
- ***A₃. Transitivité***
 - Si $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$ et
 - $B_1, B_2, \dots, B_m \rightarrow C_1, C_2, \dots, C_r$ alors
 - $A_1, A_2, \dots, A_n \rightarrow C_1, C_2, \dots, C_r$

Axiomes d'Armstrong : exemples

- A1. *Réflexivité*
 - $\{noClient\} \subseteq \{noClient, nomClient\} \Rightarrow noClient, nomClient \rightarrow noClient$
- A2. *Augmentation*
 - $noCommande, noArticle \rightarrow \text{quantité et } \{noClient\} \subseteq \{noClient, nomClient\}$
 $\Rightarrow noClient, nomClient, noCommande, noArticle \rightarrow noClient, \text{quantité}$
- A3. *Transitivité*
 - $noCommande \rightarrow noClient \text{ et } noClient \rightarrow nomClient$
 $\Rightarrow noCommande \rightarrow nomClient$

Représentation par un diagramme à bulles

- $D = \{noClient \rightarrow nomClient, noTéléphone ; noCommande \rightarrow noClient, dateCommande ; noCommande, noArticle \rightarrow quantité ; noArticle \rightarrow description, prixUnitaire\}$



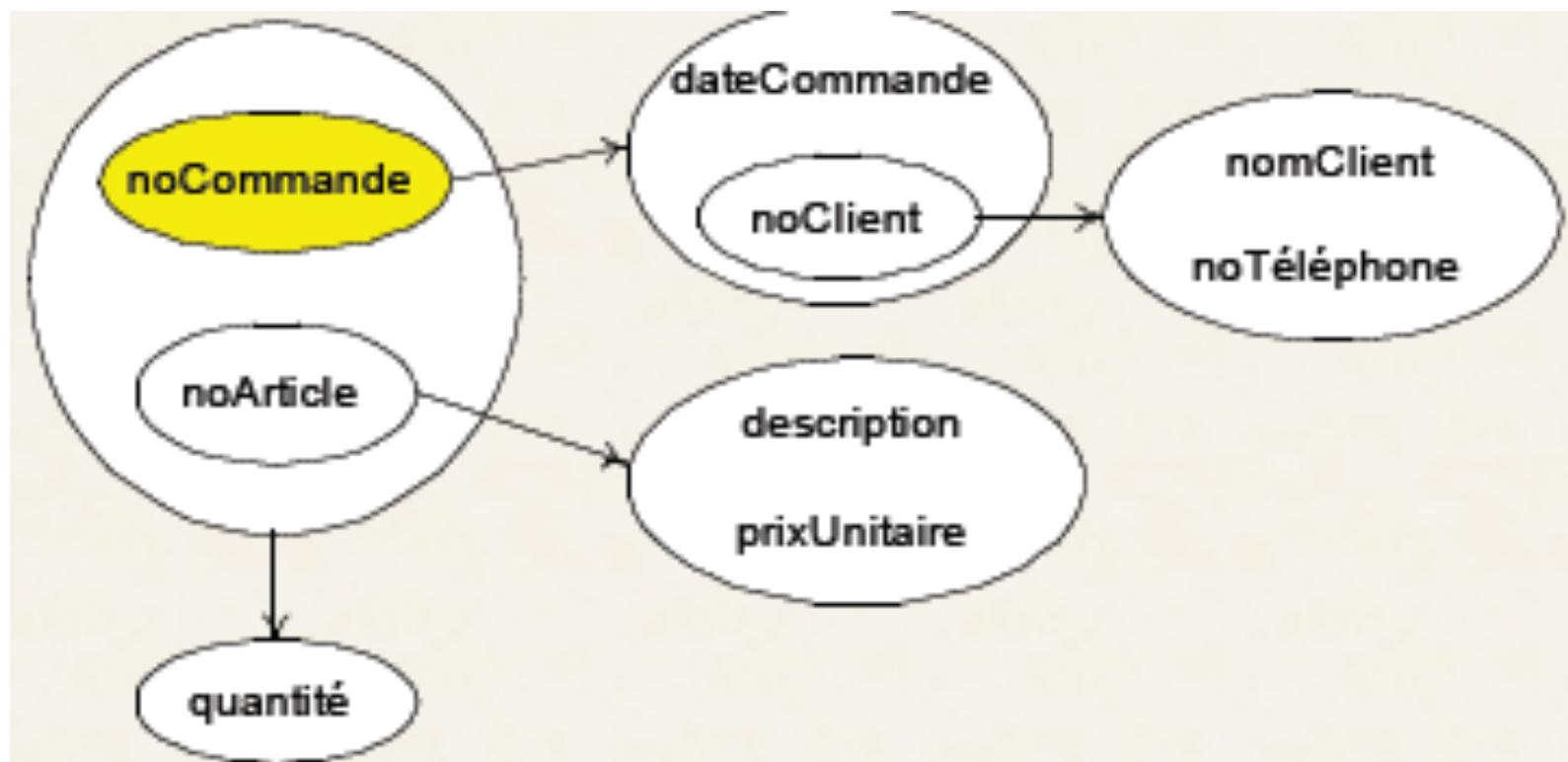
*

Fermeture d'attributs X (X^+)

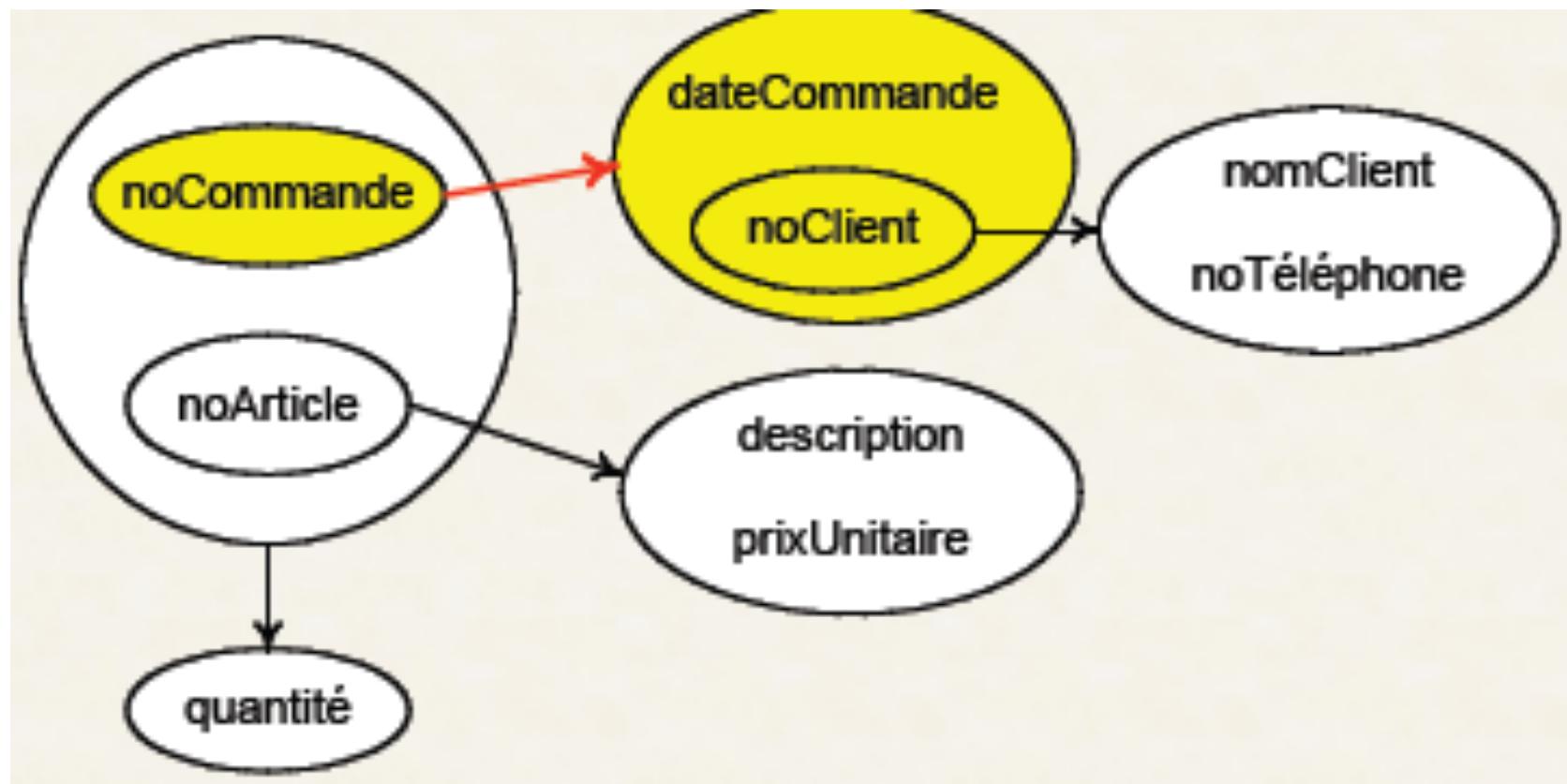
- $X^+ = \{x \mid X \rightarrow x\} = l'ensemble des attributs déterminés par X$
- Exemples
 - $\{noClient\}^+ = \{noClient, nomClient, noTéléphone\}$
 - $\{noArticle\}^+ = \{noArticle, description, prixUnitaire\}$
 - $\{prixUnitaire\}^+ = \{prixUnitaire\}$
 - $\{noCommande\}^+ = \{noCommande, noClient, dateCommande, nomClient, noTéléphone\}$
 - $\{noCommande, noArticle\}^+ = \{noClient, nomClient, noTéléphone, noCommande, dateCommande, noArticle, description, prixUnitaire, quantité\}$

Fonction Fermeture sur diagramme à bulles

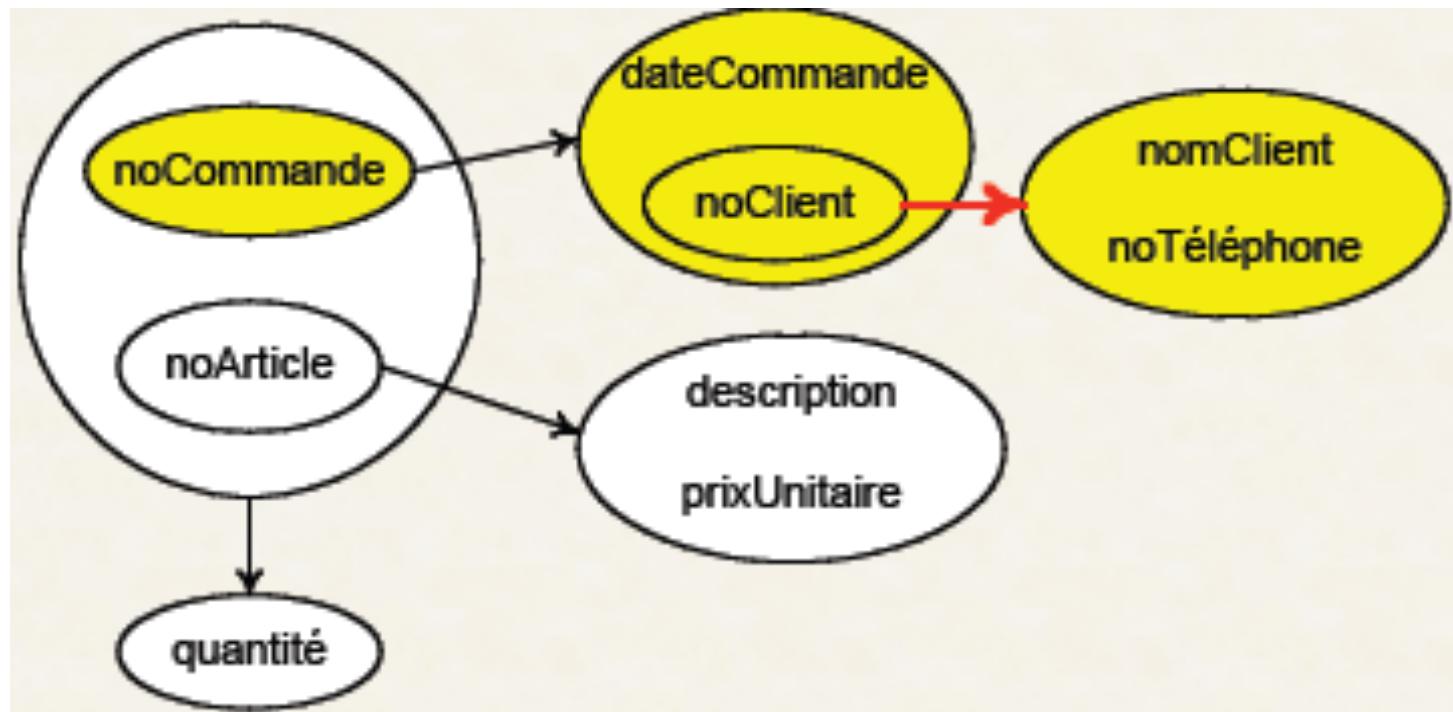
- Exemple : Fermeture(B, noCommande)



Fonction Fermeture...



Fonction Fermeture...



- $\{ noCommande \}^+ = \{ noCommande, noClient, dateCommande, nomClient, noTéléphone \}$

Fonction Dérivable ($D \models$)

$D \models d$ signifie que d est une conséquence logique de D

- $D \models X \rightarrow Y$ si et seulement si $Y \subseteq X^+$

Y est dérivable à partir de X ssi Y est un sous-ensemble de la fermeture de X

```
Fonction Dérivable(B, d)
    Entrée:
        B: le diagramme à bulle de D
        d: X → Y
    Sortie
        Vrai si D ⊨ d, faux sinon
DÉBUT
    X+ = Fermeture (B, X) ;
    Dérivable := Y ⊆ X+
FIN
```

- $D \models noCommande \rightarrow nomClient ?$
 - $\{noCommande\}^+ = \{noCommande, noClient, dateCommande, nomClient, noTéléphone\}$
 - $\{nomClient\} \subseteq \{noCommande, noClient, dateCommande, nomClient, noTéléphone\}$

Couverture minimale d'un ensemble de dépendances (D)

- Une couverture minimale, notée $\text{min}(D)$: ensemble des dépendances satisfaisant les règles suivantes:
 - 1) $\text{min}(D)$ équivalent à D
 - 2) Dépendances élémentaires
 - 3) Aucune dépendance de $\text{min}(D)$ n'est déductible des autres

*

Identification de la couverture minimale $\text{min}(D)$

- Fonction $\text{min}(D)$: en 4 étapes
 - 1) Éclater les parties de droite
 - 2) Éliminer les dépendances triviales
 - 3) Rendre les dépendances pleines, i.e. déterminant minimal
 - 4) Éliminer les dépendances déductibles des autres
 - utiliser les axiomes d'Armstrong

Fonction min(D) : étape 1

Éclater les parties droites

```
D = {noClient → noClient, nomClient, noTéléphone ;  
noCommande → noClient, noTéléphone ;  
noCommande → dateCommande ;  
noCommande, noArticle, noClient → quantité, noArticle;  
noArticle → description, prixUnitaire ;  
noCommande → nomClient }
```



```
D1 = {noClient → noClient;  
noClient → nomClient;  
noClient → noTéléphone ;  
noCommande → noClient;  
noCommande → noTéléphone ;  
noCommande → dateCommande ;  
noCommande, noArticle, noClient → quantité ;  
noCommande, noArticle, noClient → noArticle;  
noArticle → description;  
noArticle → prixUnitaire ;  
noCommande → nomClient}.
```

Fonction min(D) : étape 2

Éliminer les dépendances triviales

$$D_2 = \{ \cancel{noClient \rightarrow noClient}, \\ noClient \rightarrow nomClient, \\ noClient \rightarrow noTéléphone ; \\ noCommande \rightarrow noClient, \\ noCommande \rightarrow noTéléphone ; \\ noCommande \rightarrow dateCommande ; \\ noCommande, noArticle, noClient \rightarrow quantité ; \\ \cancel{noCommande, noArticle, noClient \rightarrow noArticle} ; \\ noArticle \rightarrow description ; \\ noArticle \rightarrow prixUnitaire ; \\ noCommande \rightarrow nomClient \}.$$

Fonction min(D) : étape 3

Rendre les dépendances pleines

$$D_3 = \{noClient \rightarrow nomClient;$$
$$noClient \rightarrow noTéléphone ;$$
$$noCommande \rightarrow noClient;$$
$$noCommande \rightarrow noTéléphone ;$$
$$noCommande \rightarrow dateCommande ;$$
$$noCommande, noArticle, \cancel{noClient} \rightarrow quantité ;$$
$$noArticle \rightarrow description;$$
$$noArticle \rightarrow prixUnitaire ;$$
$$noCommande \rightarrow nomClient\}.$$

attribut superflu

Fonction min(D) : étape 4 Éliminer les dépendances déductibles des autres

$Min(D) = \{noClient \rightarrow nomClient;$
 $noClient \rightarrow noTéléphone ;$
 $noCommande \rightarrow noClient;$
 ~~$noCommande \rightarrow noTéléphone ;$~~
 $noCommande \rightarrow dateCommande ;$
 $noCommande, noArticle \rightarrow quantité ;$
 $noArticle \rightarrow description;$
 $noArticle \rightarrow prixUnitaire ;$
 ~~$noCommande \rightarrow nomClient\}.$~~

Dépendances fonctionnelles et clés candidates

- Clé unique (**unique key**) ou superclé (**superkey**)
 - Ensemble de colonnes / Identifiant d'une ligne de la table
 - Ex: {noClient}, {noClient, nomClient}, {noClient, noTéléphone}, {nomClient, noTéléphone}, ...
- Clé candidate (**candidate key**)
 - clé unique minimale
 - Ex: {noClient}, {nomClient, noTéléphone}.
- Clé primaire (**primary key**)
 - sert de mécanisme de référence aux lignes de la table

Dépendances fonctionnelles et clés candidates... une relation étroite

- X est une clé candidate de T (U) \Leftrightarrow
 - $U = X^+$ et il n'y a pas de $Y \subset X$ tel que $U = Y^+$
i.e. aucun sous-ensemble strict de X ne peut produire U , donc X est minimal
- Pas d'algorithmes efficaces [Maier, 1983]
- Heuristique : partir des déterminants
- Visualiser par diagramme à bulles

*

Dépendances fonctionnelles et clés candidates... une relation étroite

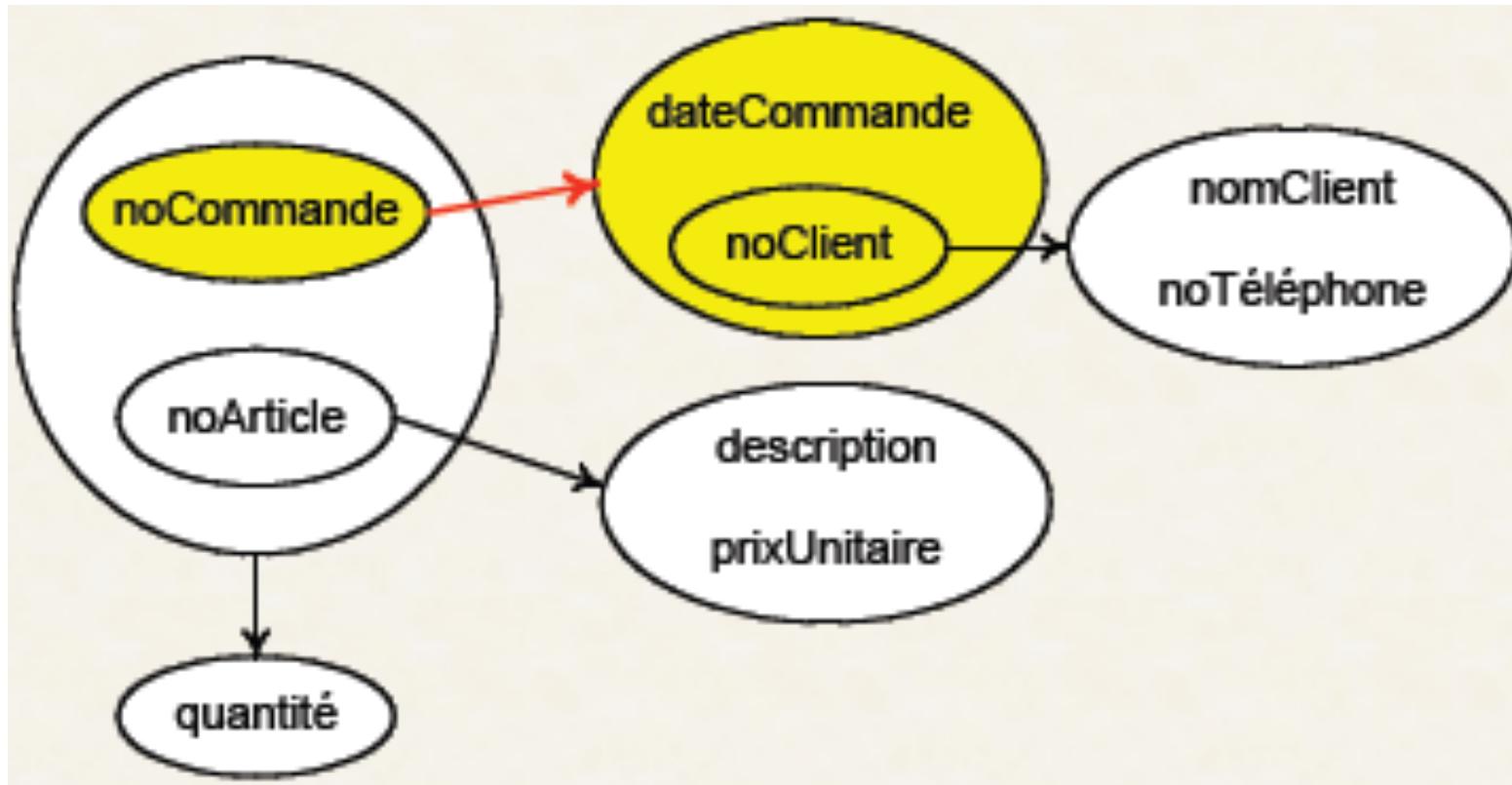
- *Exemple:*

$Min(D) = \{noClient \rightarrow nomClient;$
 $noClient \rightarrow noTéléphone ;$
 $noCommande \rightarrow noClient;$
 $noCommande \rightarrow dateCommande ;$
 $noCommande, noArticle \rightarrow quantité ;$
 $noArticle \rightarrow description;$
 $noArticle \rightarrow prixUnitaire\}.$

Dépendances fonctionnelles et clés candidates... une relation étroite

- Les fermetures des déterminants sont:
 $\{noClient\}^+ = \{noClient, nomClient, noTéléphone\}$
 $\{noCommande\}^+ = \{noCommande, noClient, dateCommande, nomClient, noTéléphone, \};$
 $\{noArticle\}^+ = \{noArticle, description, prixUnitaire\}$
 $\{noCommande, noArticle\}^+ = \{noClient, nomClient, noTéléphone, noCommande, dateCommande, noArticle, description, prixUnitaire, quantité\}$

Dépendances fonctionnelles et clés candidates... une relation étroite



- Déduction: $\{noCommande, noArticle\}^+$ est une clé candidate

Théorie de la normalisation

- Définition:
 - Résolution de problèmes de redondance de données
 - Contexte relationnel
 - Par décompositions binaires
- http://fr.wikipedia.org/wiki/Forme_normale_%28bases_de_donn%C3%A9es_relationnelles%29
- Problèmes:
 - Risque d'incohérence entre les copies des données
 - Gaspillage d'espace mémoire

Exemple de mauvaise conception

Redondance de données

Table *Vente*

noCommande	dateCommande	noClient	nomClient	noTéléphon	noArticle	description	prixUnitaire	quantité
1	01/06/2000	10	Luc Sansom	(999)999-9999	10	Cèdre en boule	10.99	10
1	01/06/2000	10	Luc Sansom	(999)999-9999	70	Herbe à puce	10.99	5
1	01/06/2000	10	Luc Sansom	(999)999-9999	90	Pommier	25.99	1
2	02/06/2000	20	D o l l a r d Tremblay	(888)888-8888	40	Epinette bleue	25.99	2
2	02/06/2000	20	D o l l a r d Tremblay	(888)888-8888	95	Génévrier	15.99	3
3	02/06/2000	10	Luc Sansom	(999)999-9999	20	Sapin	12.99	1
4	05/06/2000	10	Luc Sansom	(999)999-9999	40	Epinette bleue	25.99	1
4	05/06/2000	10	Luc Sansom	(999)999-9999	50	Chêne	22.99	1
5	09/07/2000	30	Lin Bô	(777)777-7777	70	Herbe à puce	10.99	3
5	09/07/2000	30	Lin Bô	(777)777-7777	10	Cèdre en boule	10.99	5
5	09/07/2000	30	Lin Bô	(777)777-7777	20	Sapin	12.99	5
6	09/07/2000	20	D o l l a r d Tremblay	(888)888-8888	10	Cèdre en boule	10.99	5
6	09/07/2000	20	D o l l a r d Tremblay	(888)888-8888	40	Epinette bleue	25.99	1
7	15/07/2000	40	Jean Leconte	(666)666-6666	50	Chêne	22.99	1
7	15/07/2000	40	Jean Leconte	(666)666-6666	95	Génévrier	15.99	2
8	15/07/2000	40	Jean Leconte	(666)666-6666	20	Sapin	12.99	3

- *Vente (noCl, nomCl, noTél, noCom, dateCom, noArt, desc, prixU, quantité)*

Théorie de la normalisation

- Objectifs:
 - éviter les anomalies transactionnelles pouvant découler d'une mauvaise modélisation des données;
 - éviter ou limiter les pertes de données, les incohérences au sein des données;
 - éviter les anomalies de lecture, d'écriture, la redondance des données et la contre performance.
- Avantage:
 - vérifier la robustesse de la conception des données pour améliorer la modélisation pour une meilleure représentation
 - faciliter la mémorisation des données en évitant la redondance et les problèmes sous-jacents de mise à jour ou de cohérence.

**

Théorie de la normalisation

- *Les différentes formes normales:*
 - (1FN et 2FN) ...sans intérêt pour relationnel
 - dépendances fonctionnelles (3FN et FNBC)
 - dépendances multivaluées (4FN)
 - dépendances de jointures (5FN)
 - dépendances généralisées
 - contraintes de domaine (FNDC)

Normalisation

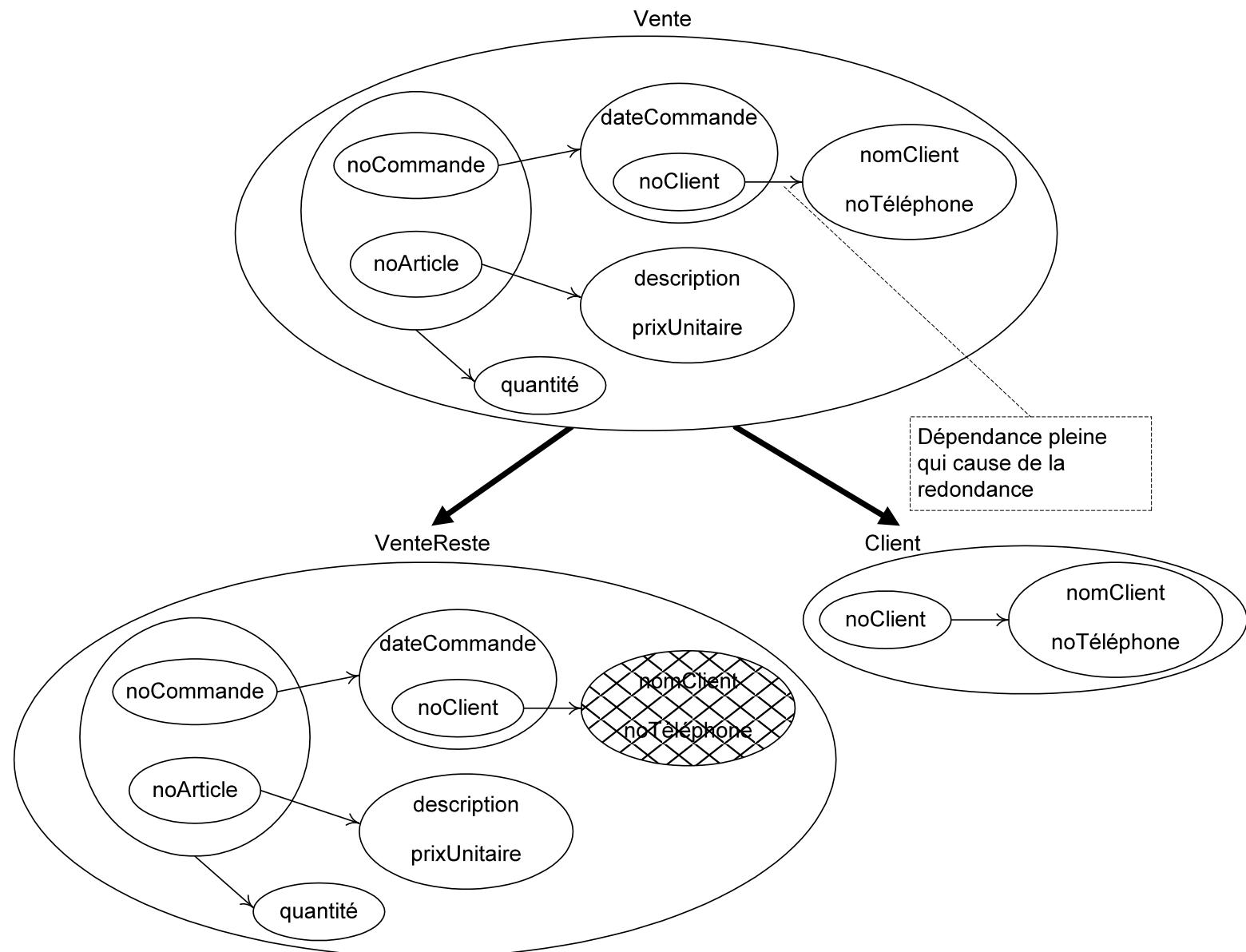
- Rappel:
– Résolution de problèmes de redondance de données

noClient n'est pas une clé candidate

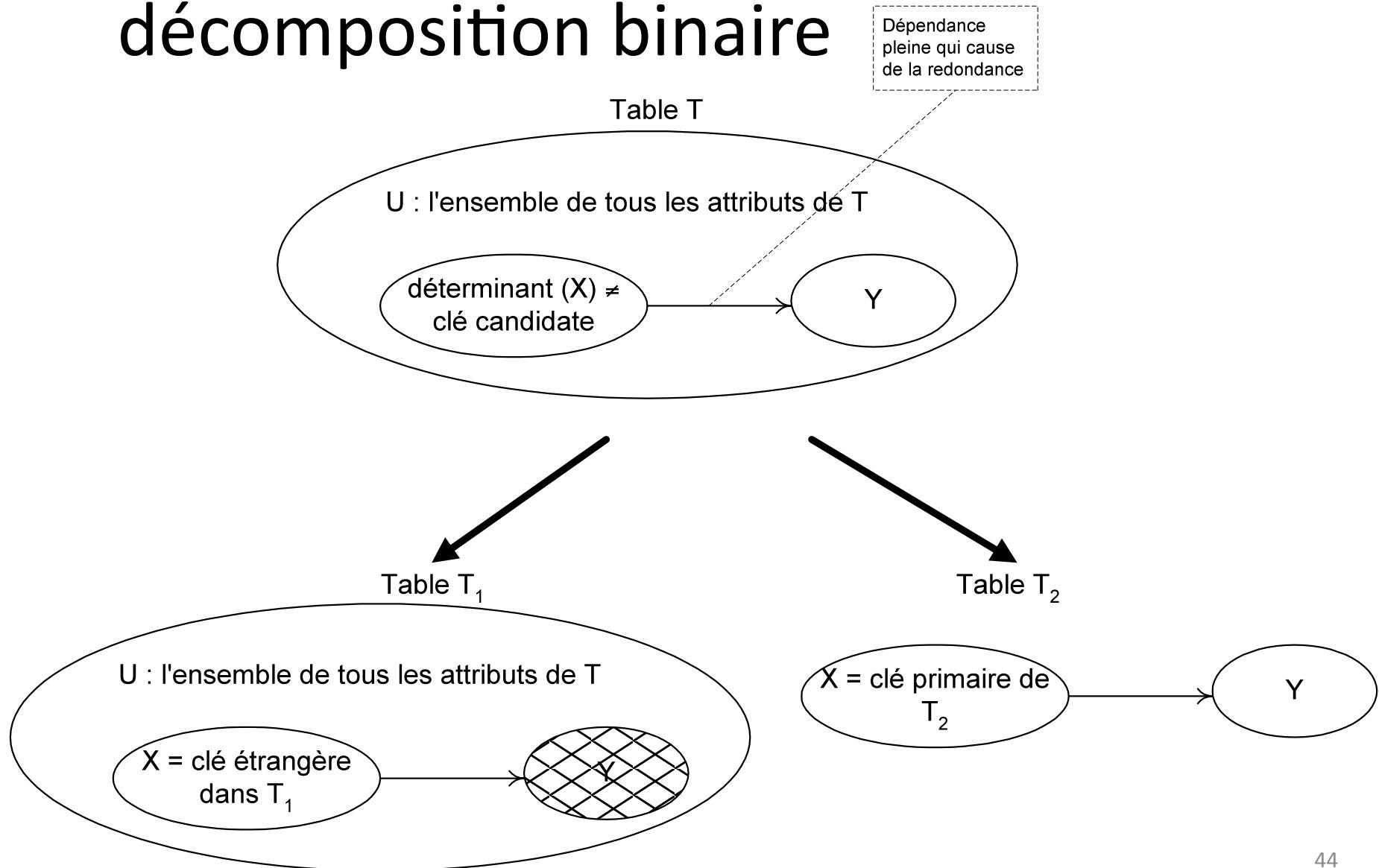
noClient->nomClient, noTéléphone

Table Vente									
noCommande	dateCommande	noChent	nomChent	noTéléphone	noArticle	description	prixUnitaire	quantité	
1	01/06/2000	10	Luc Sansom	(999)999-9999	10	Cèdre en boule	10.99	10	
1	01/06/2000	10	Luc Sansom	(999)999-9999	70	Herbe à puce	10.99	5	
1	01/06/2000	10	Luc Sansom	(999)999-9999	90	Pommier	25.99	1	
2	02/06/2000	20	D o l l a r d Tremblay	(888)888-8888	40	Epinette bleue	25.99	2	
2	02/06/2000	20	D o l l a r d Tremblay	(888)888-8888	95	Génévrier	15.99	3	
3	02/06/2000	10	Luc Sansom	(999)999-9999	20	Sapin	12.99	1	
4	05/06/2000	10	Luc Sansom	(999)999-9999	40	Epinette bleue	25.99	1	
4	05/06/2000	10	Luc Sansom	(999)999-9999	50	Chêne	22.99	1	
5	09/07/2000	30	Lin Bô	(777)777-7777	70	Herbe à puce	10.99	3	
5	09/07/2000	30	Lin Bô	(777)777-7777	10	Cèdre en boule	10.99	5	
5	09/07/2000	30	Lin Bô	(777)777-7777	20	Sapin	12.99	5	
6	09/07/2000	20	D o l l a r d Tremblay	(888)888-8888	10	Cèdre en boule	10.99	5	

Normalisation



Principe général de décomposition binaire



Forme normale Boyce-Codd (FNBC)

- Table en FNBC
 - Pour toute dépendance fonctionnelle non triviale et pleine, $X \rightarrow^T Y$, le déterminant X est une clé candidate de T
 - OU
 - Pour toute dépendance fonctionnelle non triviale, $X \rightarrow^T Y$, le déterminant X est une superclé de T
 - Rappel : clé unique (unique key) ou superclé (superkey)
 - X est une clé unique (ou superclé) de T si deux lignes de T ne peuvent avoir les mêmes valeurs pour toutes les colonnes de X
- Schéma relationnel est en FNBC
 - si toutes les tables du schéma sont en FNBC

Algorithme de décomposition en FNBC

Procédure *DécompositionFNBC* (*T*, *S*)

Entrée:

T : une table avec ses dépendances fonctionnelles élémentaires

Sortie

S : un schéma relationnel pour *T* en FNBC

DÉBUT

Ajouter *T* à *S*

TANT QU'il y a une table *T* dans *S* qui n'est pas en FNBC

 Décomposer *T* selon une dépendance pleine $X \rightarrow^T Y$ qui ne respecte
 pas la condition de FNBC

 et remplacer *T* par $T_1(U-Y)$ et $T_2(X \cup Y)$ dans *S*

FIN TANT QUE

FIN

Décomposition de Vente

Vente (noCl, nomCl, noTél, noCom, dateCom, noArt, desc, prixU, quantité)

- Les dépendances pleines:

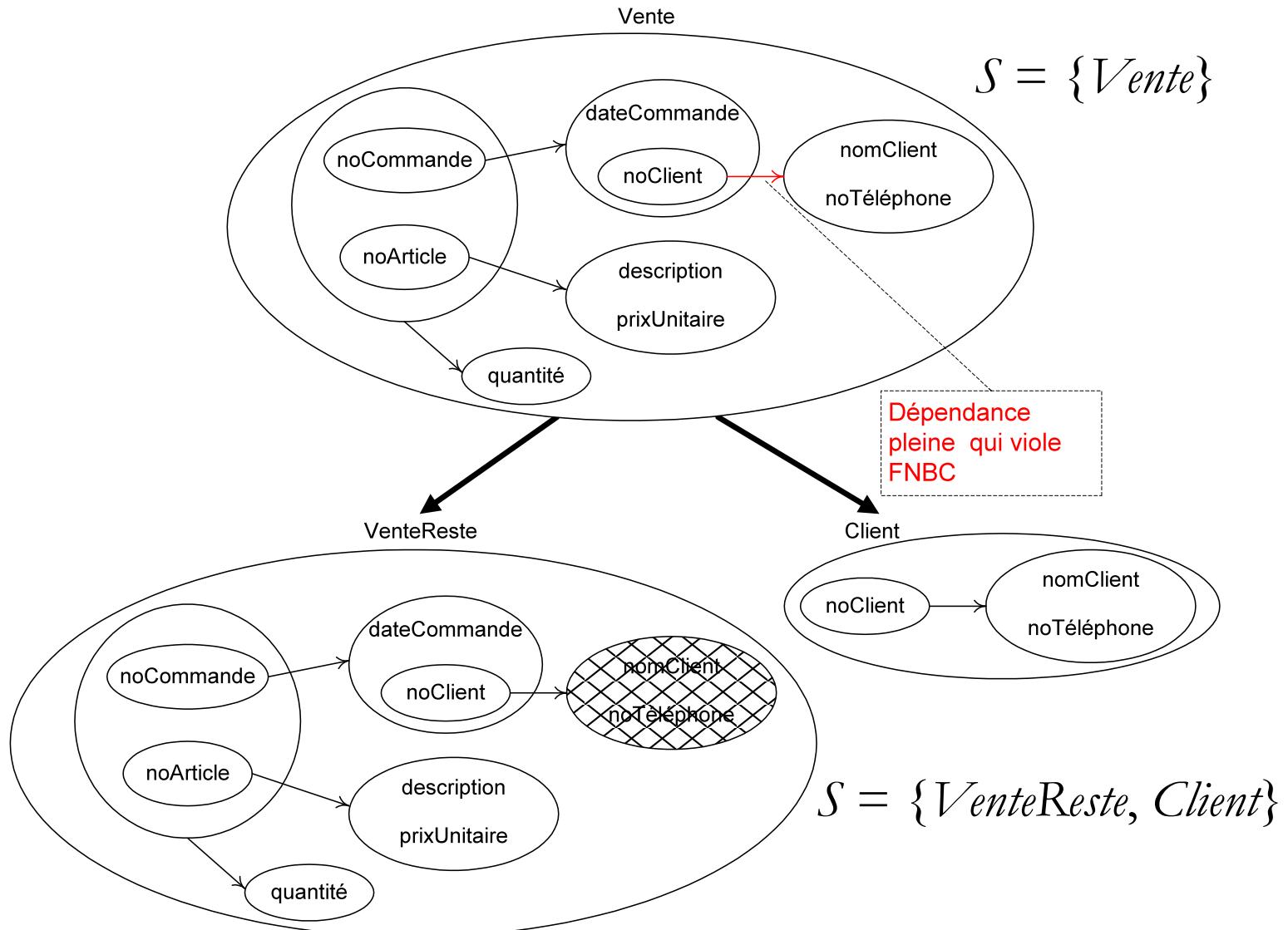
$D = \{noClient \rightarrow nomClient, noTelephone;$
 $noCommande \rightarrow noClient, dateCommande;$
 $noCommande, noArticle \rightarrow quantité;$
 $noArticle \rightarrow description, prixUnitaire\}$

Initialisation: $S = \{Vente\}$

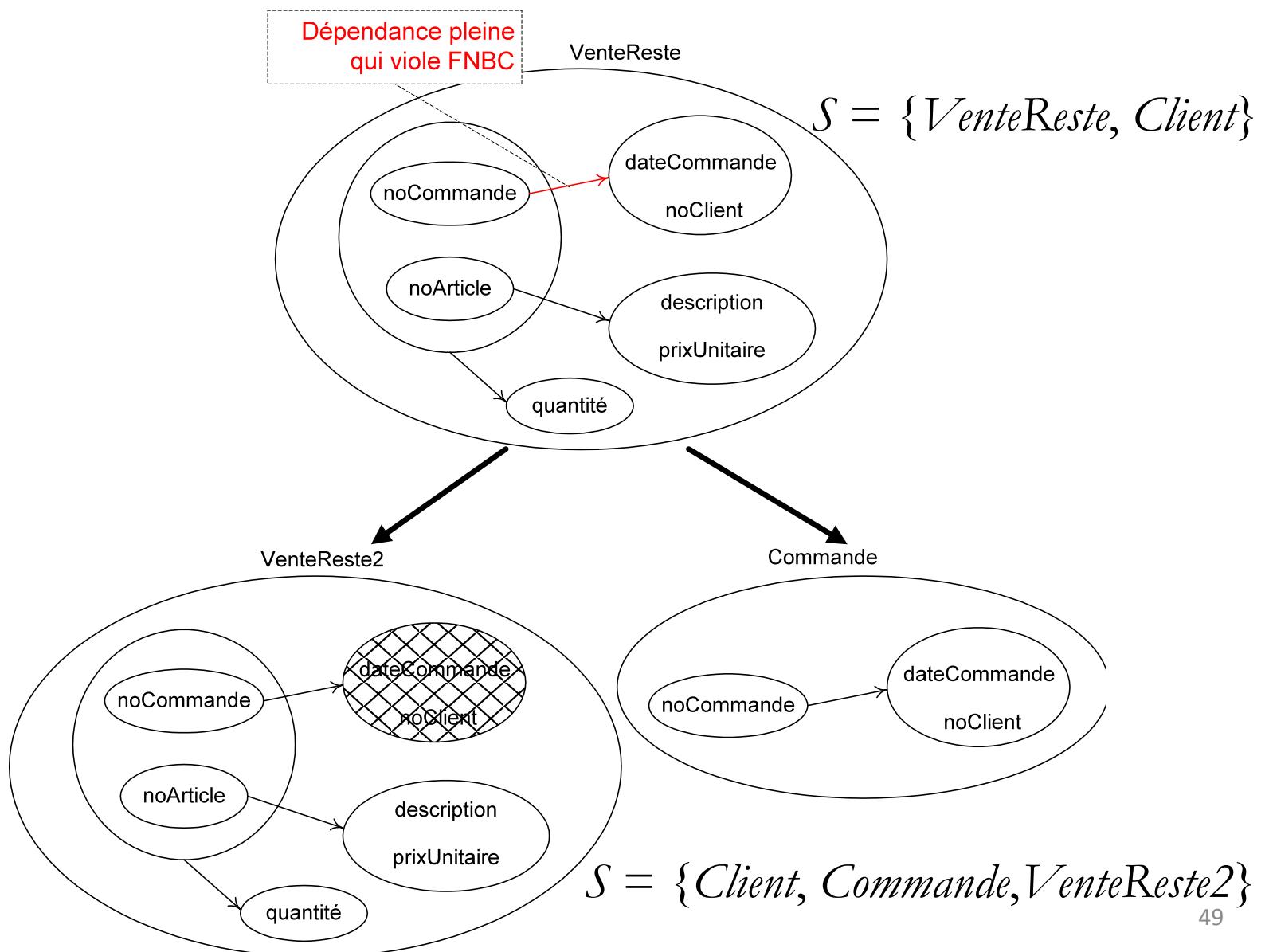
Clé candidate={noCommande, noArticle}

Décomposition de Vente

Vente (noCl, nomCl, noTél, noCom, dateCom, noArt, desc, prixU, quantité)

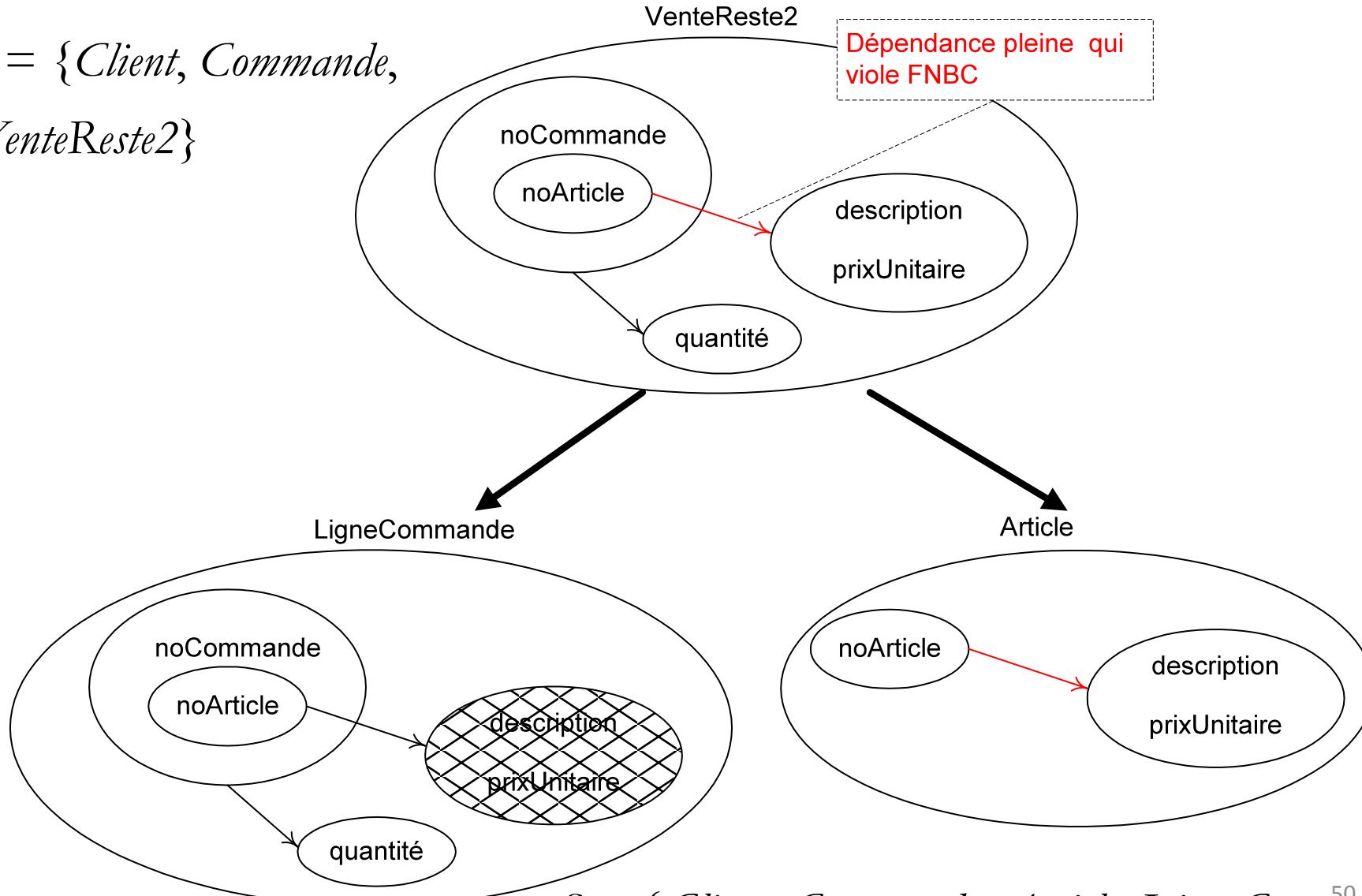


Décomposition de Vente (suite)



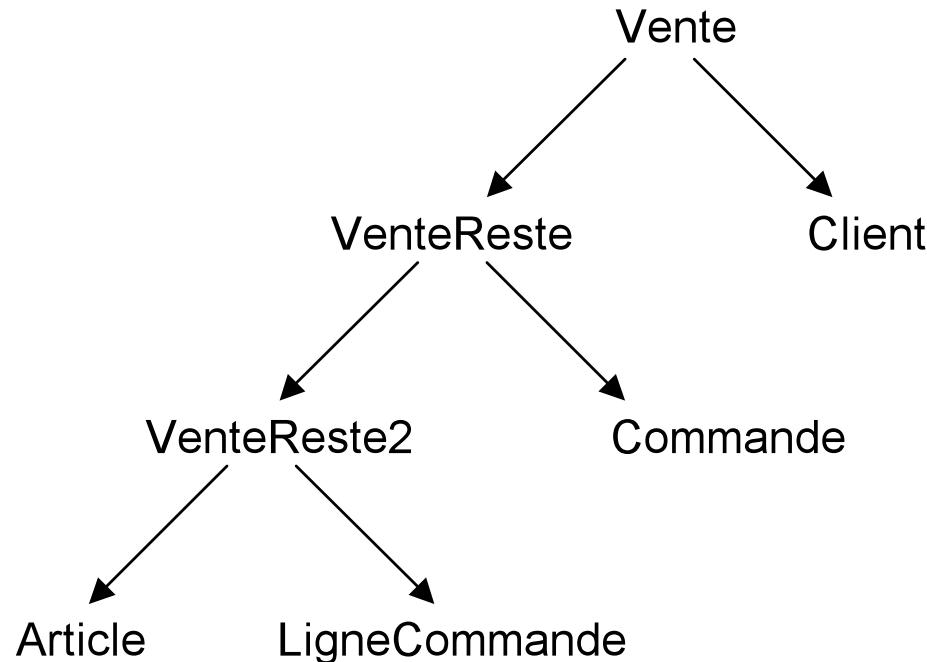
Décomposition de Vente (suite)

$S = \{Client, Commande,$
 $VenteReste2\}$



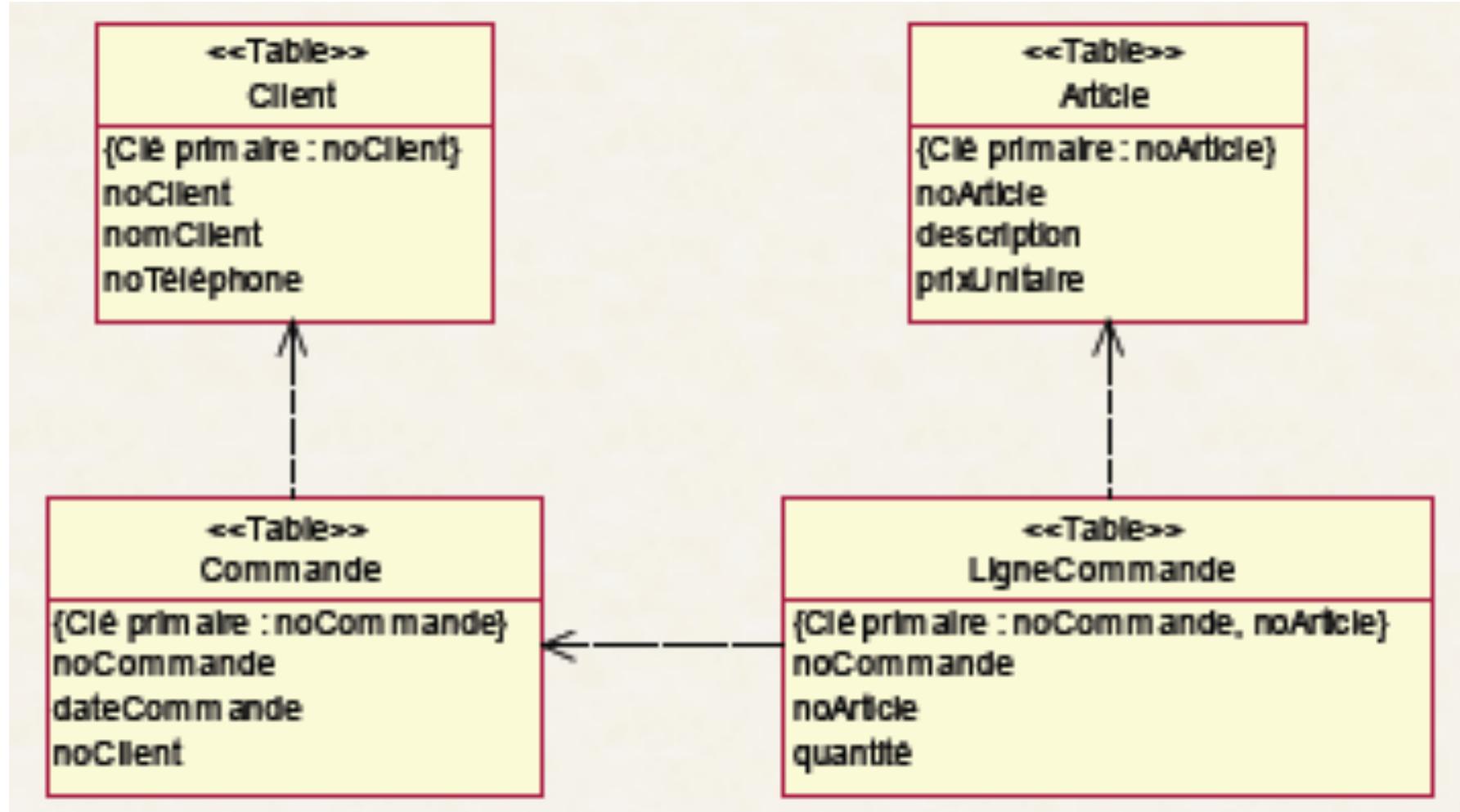
$S = \{Client, Commande, Article, LigneCommande\}$

Schéma final (S)



- *Client (noClient, nomClient, noTéléphone)*
- *Commande(noCommande, dateCommande, noClient)*
- *LigneCommande(noCommande, noArticle, quantité)*
- *Article(noArticle, description, prixUnitaire)*

Résultat en UML



Troisième forme normale

- FNBC ne préserve pas toujours DF

Table <i>Adresse</i>		
rue	ville	codePostal
Chambly	Longueuil	J4G 2L7
Chambly	St-Hubert	J3S 2R4
Taschereau	Longueuil	J4G 2L7
Taschereau	St-Hubert	J3S 4T8

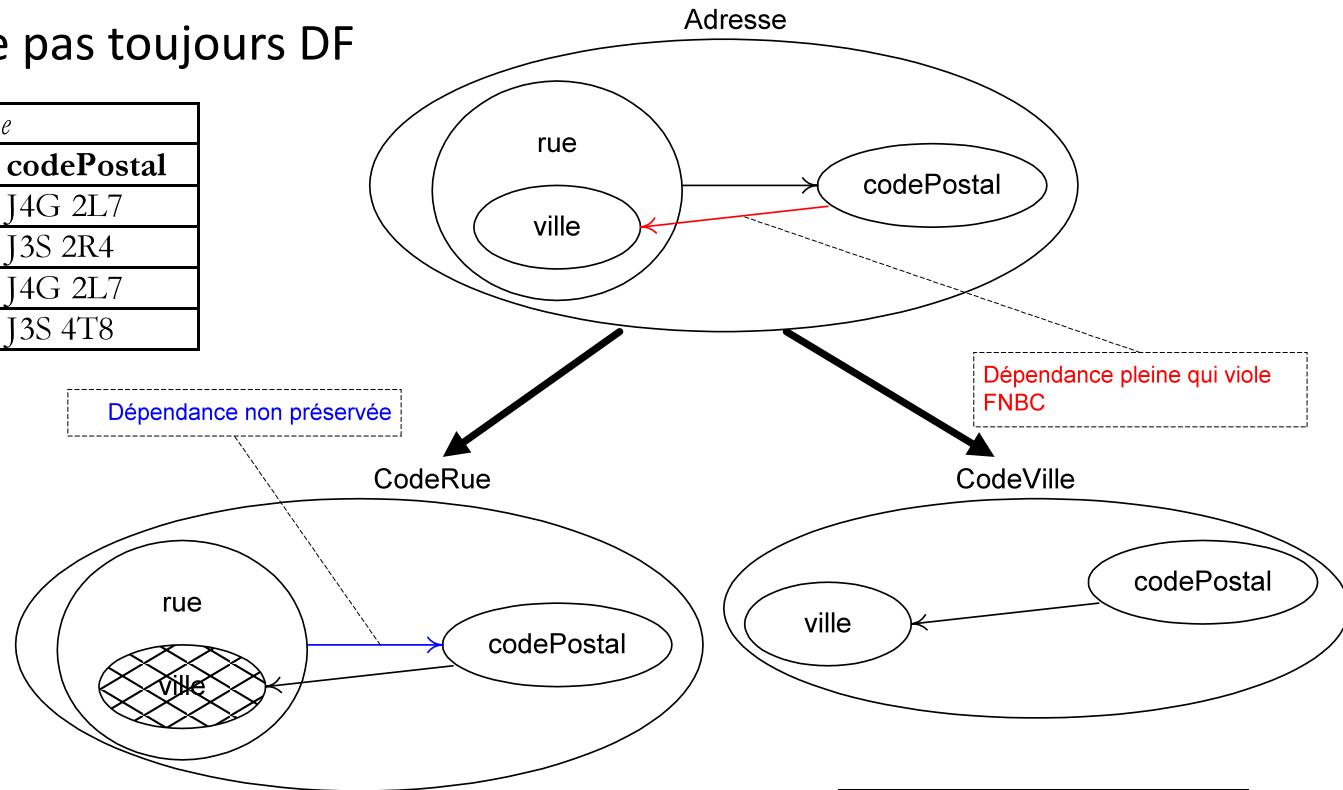


Table <i>CodeRue</i>	
codePostal	rue
J4G 2L7	Chambly
J3S 2R4	Chambly
J4G 2L7	Taschereau
J3S 4T8	Taschereau

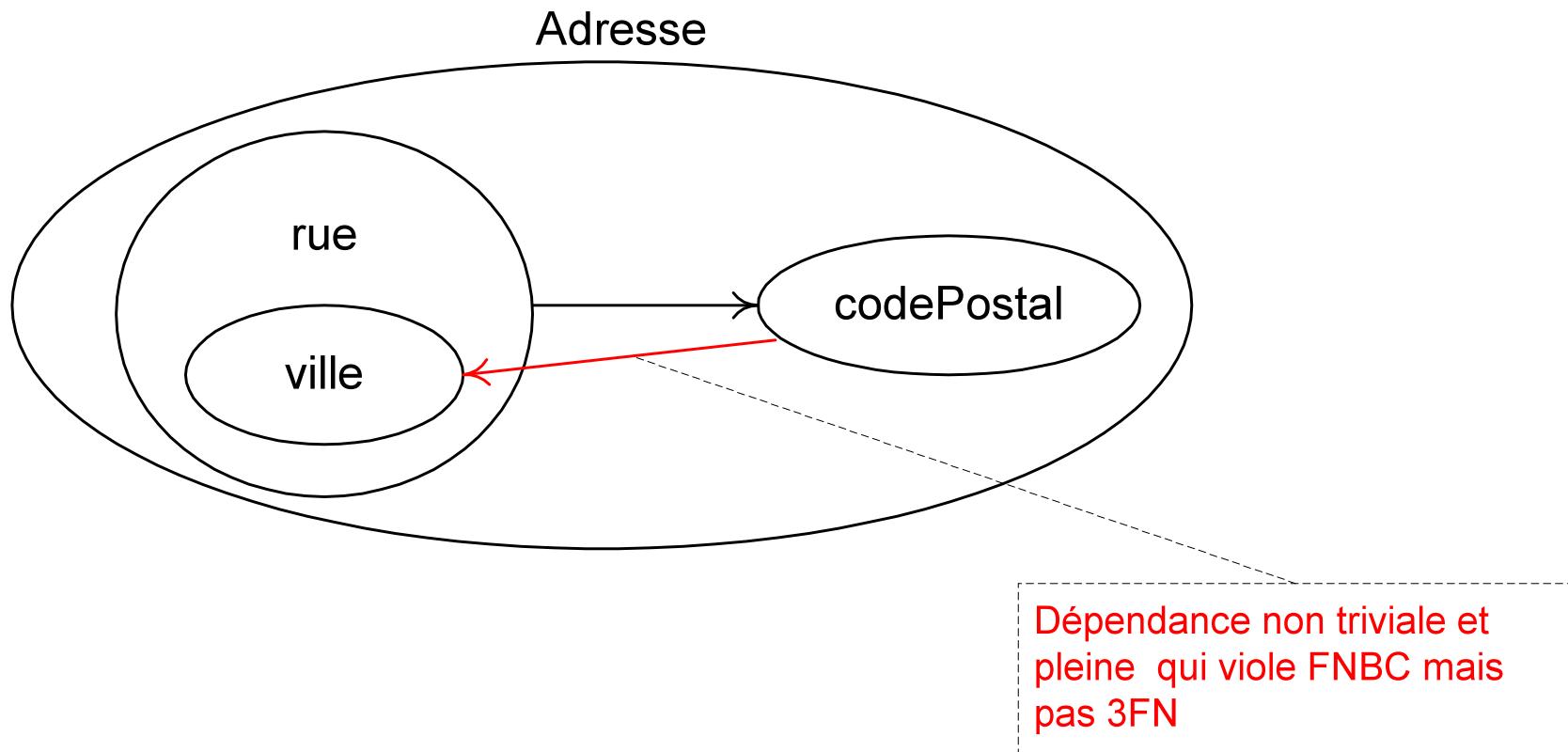
Table <i>CodeVille</i>	
codePostal	ville
J4G 2L7	Longueuil
J3S 2R4	St-Hubert
J3S 4T8	St-Hubert

Troisième forme normale (3FN, third normal form)

- Pour toute dépendance fonctionnelle non triviale et pleine, $X \rightarrow^T Y$, le déterminant X est une clé candidate de T **ou Y fait partie d'une clé candidate**
- ex: codePostal \rightarrow ville $\in \{\text{rue}, \text{ville}\}$

Adresse est en 3FN

- *ville* est première
 - fait partie de la clé candidate $\{rue, ville\}$



3 FN

- **Proposition.** Il y a toujours une décomposition en 3FN qui préserve les dépendances.
- **Proposition.** FNBC \Rightarrow 3FN

Algorithme de synthèse pour la 3FN

Procédure *Synthèse3FN* (*T, S*)

Entrée:

T : une table avec ses dépendances fonctionnelles *D*

Sortie

S : un schéma relationnel pour *T* en 3FN
qui préservent les dépendances

DÉBUT

Calculer $\min(D)$;

S := \emptyset ;

Former la dépendance $X \rightarrow a_1, a_2, \dots, a_n$ avec l'ensemble des dépendances
de $\min(D)$ de la forme $X \rightarrow a_i$, $i = 1, \dots, n$

POUR chaque dépendance $X \rightarrow Y = a_1, a_2, \dots, a_n$ dans $\min(D)$

Ajouter une table $T(X \cup Y)$ à *S*; *X* devient une clé candidate de *T*

FIN POUR;

Eliminer les tables dont le schéma est inclus dans une autre table de *S*;
SI aucune table de *S* ne contient une clé candidate de *T*

Ajouter une table avec les colonnes d'une clé candidate

FIN SI;

Fusionner les tables dont les clés candidates se déterminent
mutuellement

FIN

Exemple avec Ventes

```
D = {noClient → noClient, nomClient, noTéléphone ;  
noCommande → noClient, noTéléphone ;  
noCommande → dateCommande ;  
noCommande, noArticle, noClient → quantité, noArticle;  
noArticle → description, prixUnitaire ;  
noCommande → nomClient }
```



```
min(D) = {noClient → nomClient;  
noClient → noTéléphone ;  
noCommande → noClient;  
noCommande → dateCommande ;  
noCommande, noArticle → quantité ;  
noArticle → description;  
noArticle → prixUnitaire }.
```

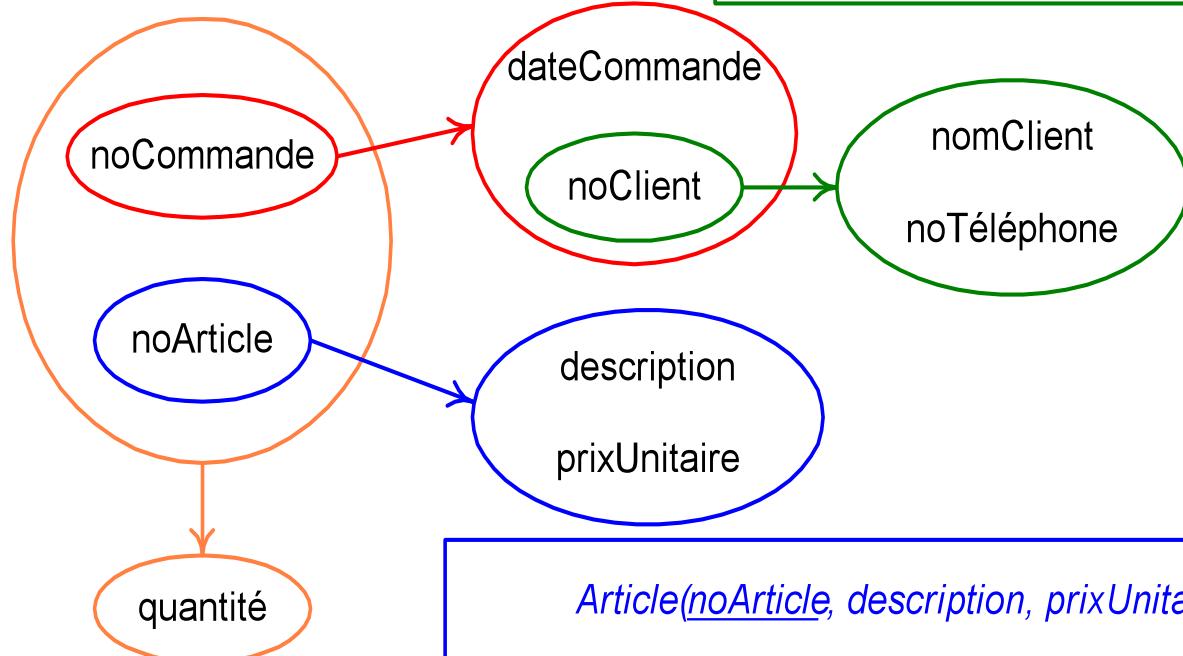
Regroupement des dépendances de
même déterminant

```
{noClient → nomClient, noTéléphone ;  
noCommande → noClient, dateCommande ;  
noCommande, noArticle, → quantité ;  
noArticle → description, prixUnitaire}
```

Par le diagramme à bulles

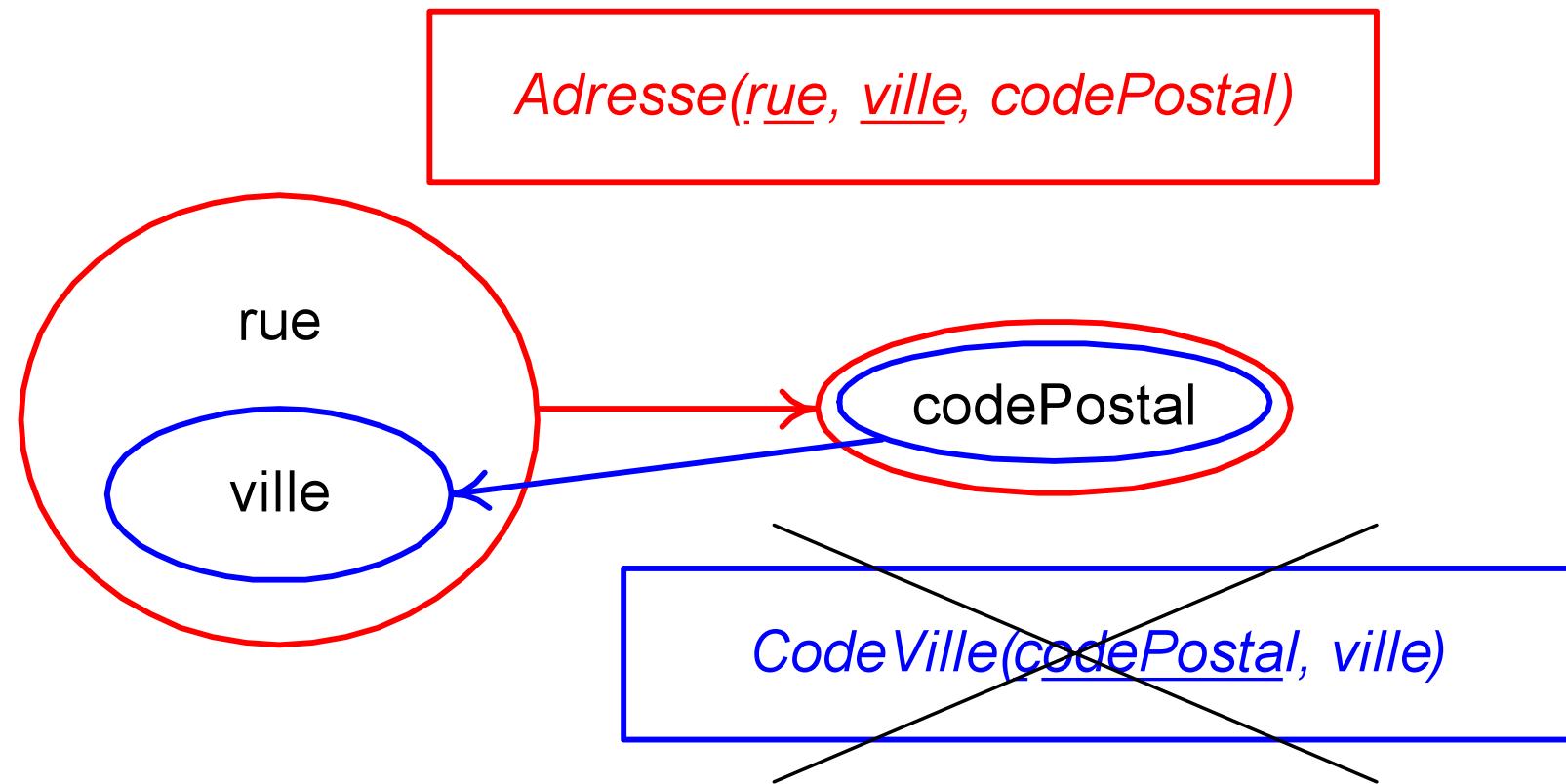
Commande(noCommande, dateCommande, noClient)

Client (noClient, nomClient, noTéléphone)



LigneCommande(noCommande, noArticle, quantité)

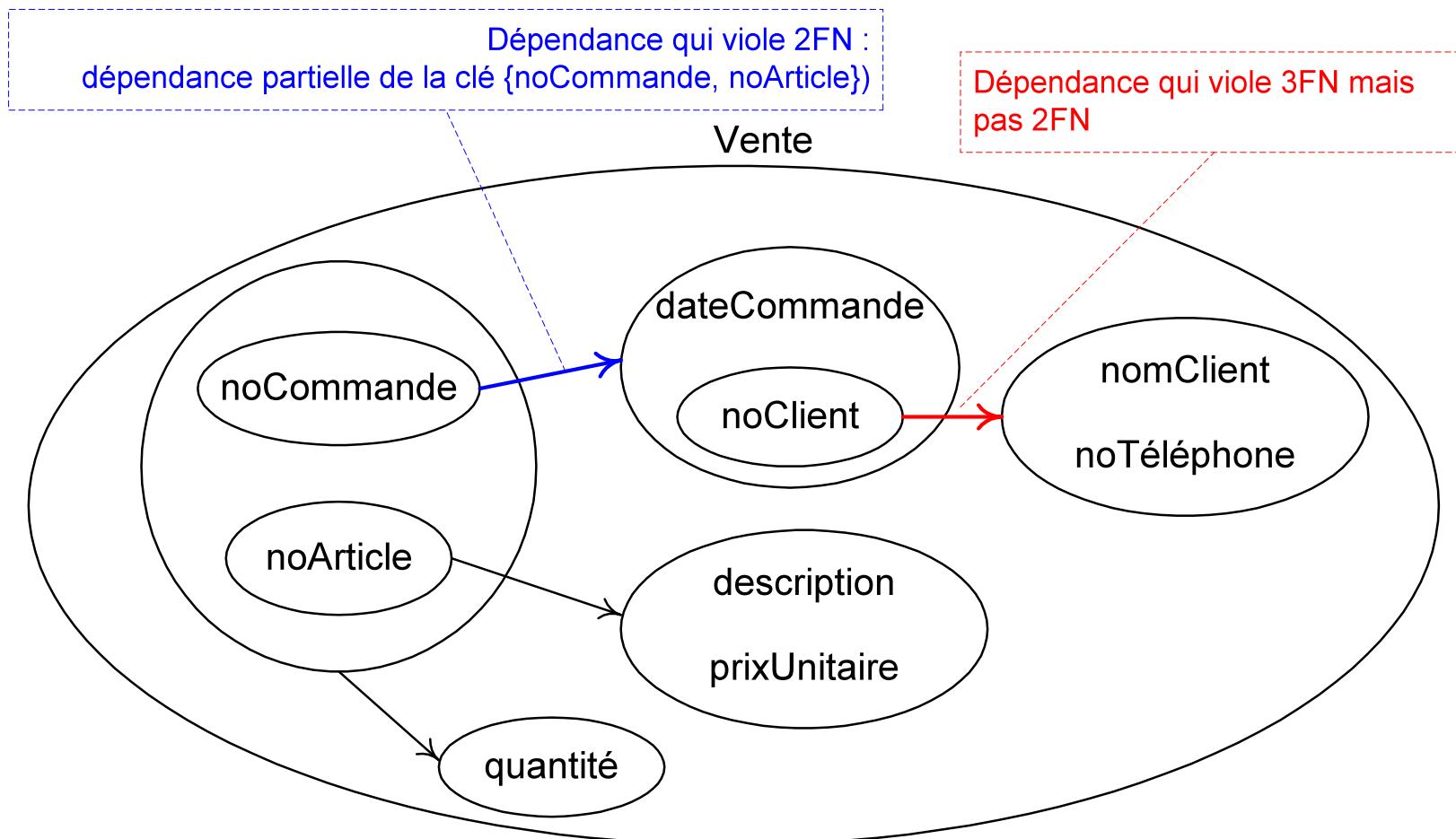
Exemple avec Adresse



Résultat en 3FN mais pas en FNBC

Vue d'ensemble des formes normales

- 2FN : interdit dépendance partielles d 'une clé ($X \subset$ clé candidate)



Vue d'ensemble des formes normales

- 1FN : la table ne doit pas comporter des groupes répétitifs
 - chaque ligne de la table doit avoir une seule valeur pour chaque colonne
 - vrai pour le modèle relationnel

1FN : valeurs atomiques

Table Vente avec table encastrée (non en 1FN)						
noCommande	dateCommande	noClient	nomClient	noTéléphone	lignesDeLaCommande (table encastrée)	
					noArticle	description
1	01/06/2000	10	Hugh Paycheck	(999)999-9999	10	Cèdre en boule
					70	Herbe à poee
					90	Pommier
2	02/06/2000	20	Dollard Cash	(888)888-8888	40	Epinette bleue
					95	Génévrier
					20	Sapin
3	02/06/2000	10	Hugh Paycheck	(999)999-9999	40	Epinette bleue
					50	Chêne
					70	Herbe à poee
4	05/06/2000	10	Hugh Paycheck	(999)999-9999	20	Cèdre en boule
					50	Sapin
					70	Génévrier
5	09/07/2000	30	Ye San Le Sou	(777)777-7777	20	Chêne
					70	Sapin
					10	Génévrier
6	09/07/2000	20	Dollard Cash	(888)888-8888	20	Herbe à poee
					40	Cèdre en boule
					50	Epinette bleue
7	15/07/2000	40	Le Comte Hazeck	(666)666-6666	50	Chêne
					95	Sapin
					20	Génévrier
8	15/07/2000	40	Le Comte Hazeck	(666)666-6666	20	Herbe à poee
					70	Cèdre en boule

Modèle relationnel étendu ou NF² (Non First Normal Form)

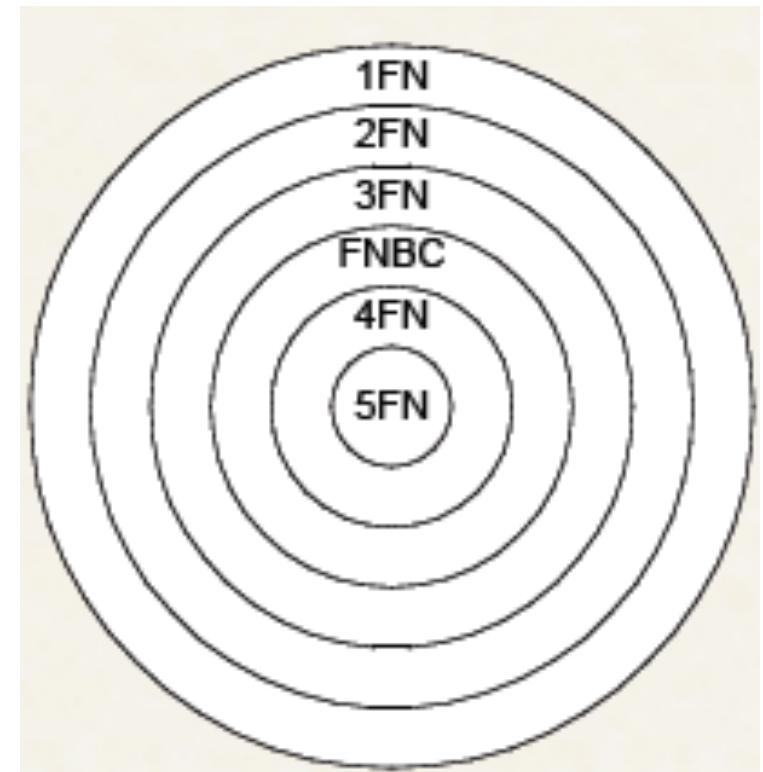
- Colonnes à valeurs complexes
- Table enchaînée
 - Oracle8
 - NESTED TABLE
- Valeur composée
 - Oracle8/SQL:1999
 - CREATE TYPE *nomType* AS...

Modèle relationnel étendu ou NF² (Non First Normal Form)

- Extensions objet
 - Table d'objets
 - Oracle8/SQL:1999
 - CREATE TABLE *nom* OF *nomType*
 - Colonne qui réfère à un objet
 - Oracle8/SQL:1999
 - REF

Restrictions de plus en plus fortes

- $5FN \Rightarrow 4FN \Rightarrow FNBC \Rightarrow 3FN \Rightarrow 2FN \Rightarrow 1FN$
- 2FN: dépendances élémentaire dont le déterminant n'est pas une clé candidate
- 3FN interdit les déterminants non clé
- FNBC prend en compte toutes les dépendances



Quatrième forme normale et dépendances multivaluées

- *Film* est en FNBC mais pas en 4FN
- Redondance due à dépendance multivaluée
 - $titre \rightarrow\rightarrow nomProducteur$
 - $titre \rightarrow\rightarrow nomActeur$

Table <i>Film</i>		
titre	nomProducteur	nomActeur
La vie est belle	Elda Ferri	Roberto Benigni
La vie est belle	Elda Ferri	Nicoletta Braschi
La vie est belle	Elda Ferri	Giorgio Cantarini
La vie est belle	Gianluigi Braschi	Roberto Benigni
La vie est belle	Gianluigi Braschi	Nicoletta Braschi
La vie est belle	Gianluigi Braschi	Giorgio Cantarini
Patch Adams	Barry Kemp	Robin Williams
Patch Adams	Barry Kemp	Monica Potter
Patch Adams	Michael Farrell	Robin Williams
Patch Adams	Michael Farrell	Monica Potter
Patch Adams	Marvin Minoff	Robin Williams
Patch Adams	Marvin Minoff	Monica Potter

$A \rightarrow\rightarrow^T B$ et $A \rightarrow\rightarrow^T C$ mais $B \neg\leftrightarrow^T C$

Dépendance multivaluée (multivalued dependency)

- $A_1, A_2, \dots, A_n \rightarrow\rightarrow^T B_1, B_2, \dots, B_m$
 - à chacune de valeurs de A_1, A_2, \dots, A_n est associé un ensemble de valeurs de B_1, B_2, \dots, B_m sans relation directe avec les autres colonnes C_1, C_2, \dots, C_p de la table
- $A_1, A_2, \dots, A_n \rightarrow\rightarrow^T C_1, C_2, \dots, C_p$
 - vient par paire

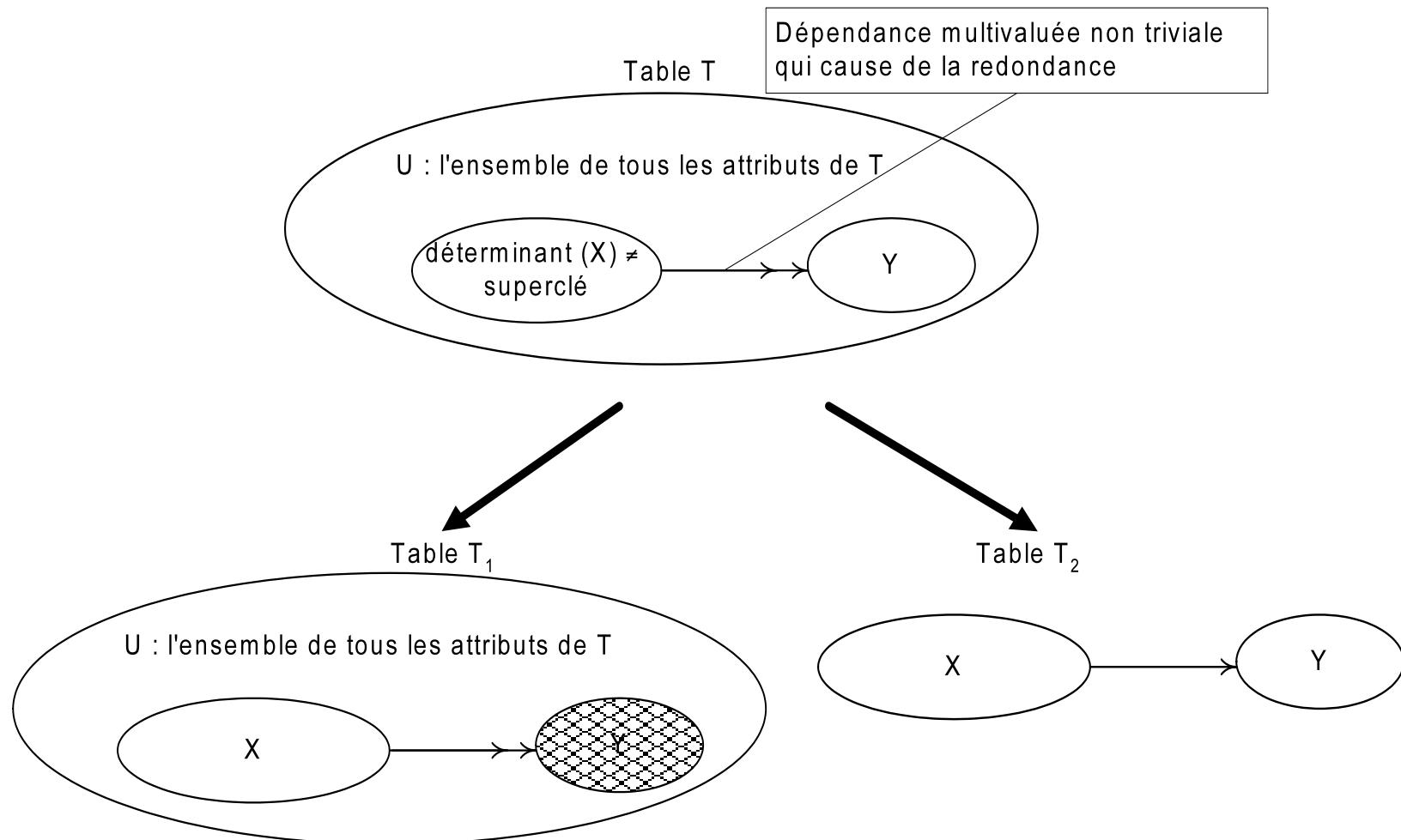
Définition formelle (règle du complément)

- $(a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m, c_1, c_2, \dots, c_p) \in T \Rightarrow (a_1, a_2, \dots, a_n, b'_1, b'_2, \dots, b'_m, c'_1, c'_2, \dots, c'_p) \in T$
- $('La vie est belle', 'Elda Ferri', 'Robert Benigni') \in \text{Film} \Rightarrow ('La vie est belle', 'Gianluigi Braschi', 'Nicoletta Braschi') \in \text{Film}$
- $A \rightarrow TB$ et $A \rightarrow TC$ mais $B \neg\leftrightarrow TC \Rightarrow A \rightarrow \rightarrow TB$ et $A \rightarrow \rightarrow TC$

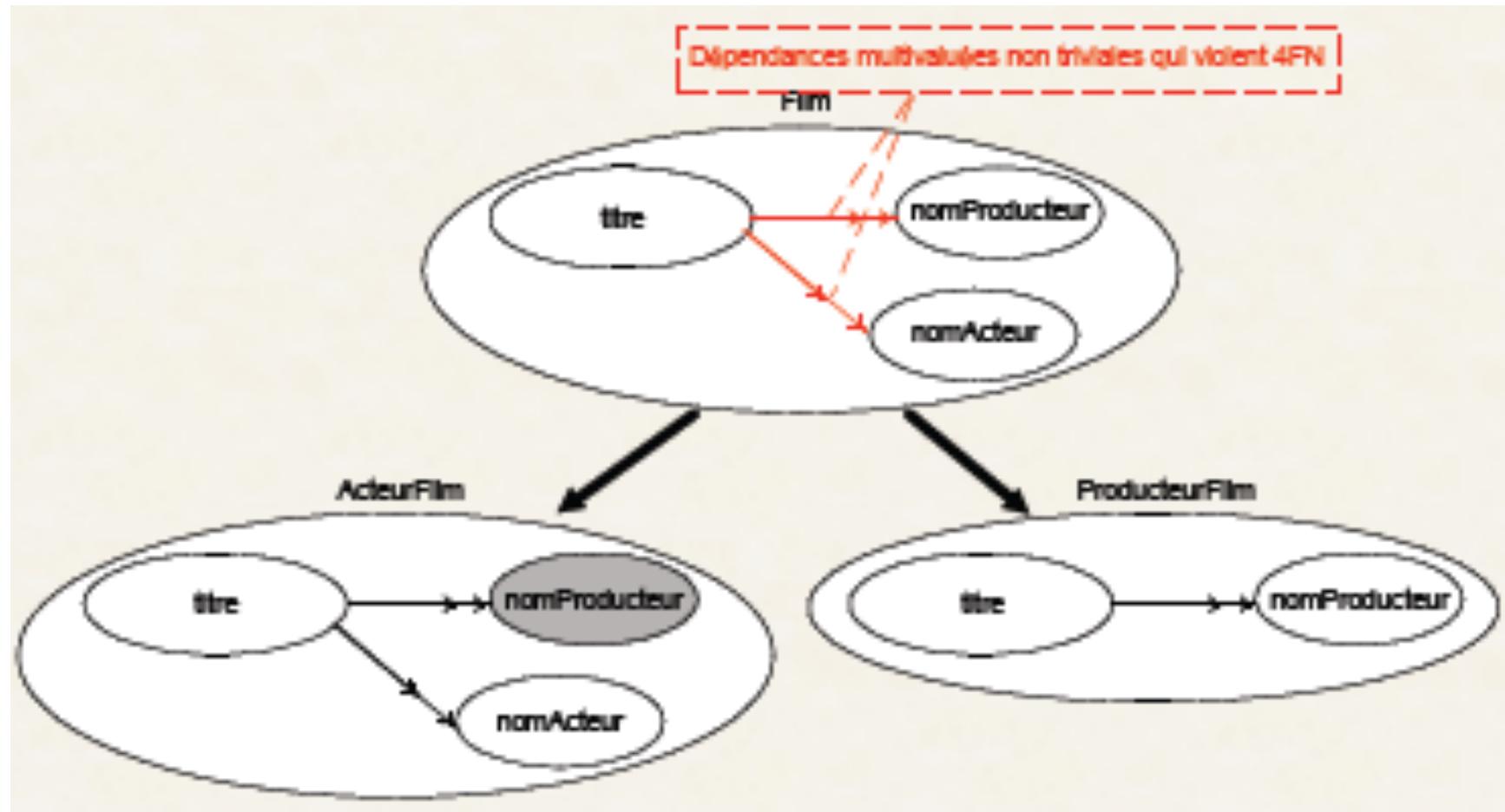
Dépendances multivaluées

- **Proposition (règle du complément)**
 - $A_1, A_2, \dots, A_n \rightarrow\rightarrow_T B_1, B_2, \dots, B_m \Rightarrow$
 $A_1, A_2, \dots, A_n \rightarrow\rightarrow_T C_1, C_2, \dots, C_p$
 - C_1, C_2, \dots, C_p correspond aux autres colonnes de T
- **Proposition (DF cas particulier de DMV)**
 - $A_1, A_2, \dots, A_n \rightarrow_T B_1, B_2, \dots, B_m \Rightarrow$
 $A_1, A_2, \dots, A_n \rightarrow\rightarrow_T B_1, B_2, \dots, B_m$
($DF = DMV$ où $|B|=1$)

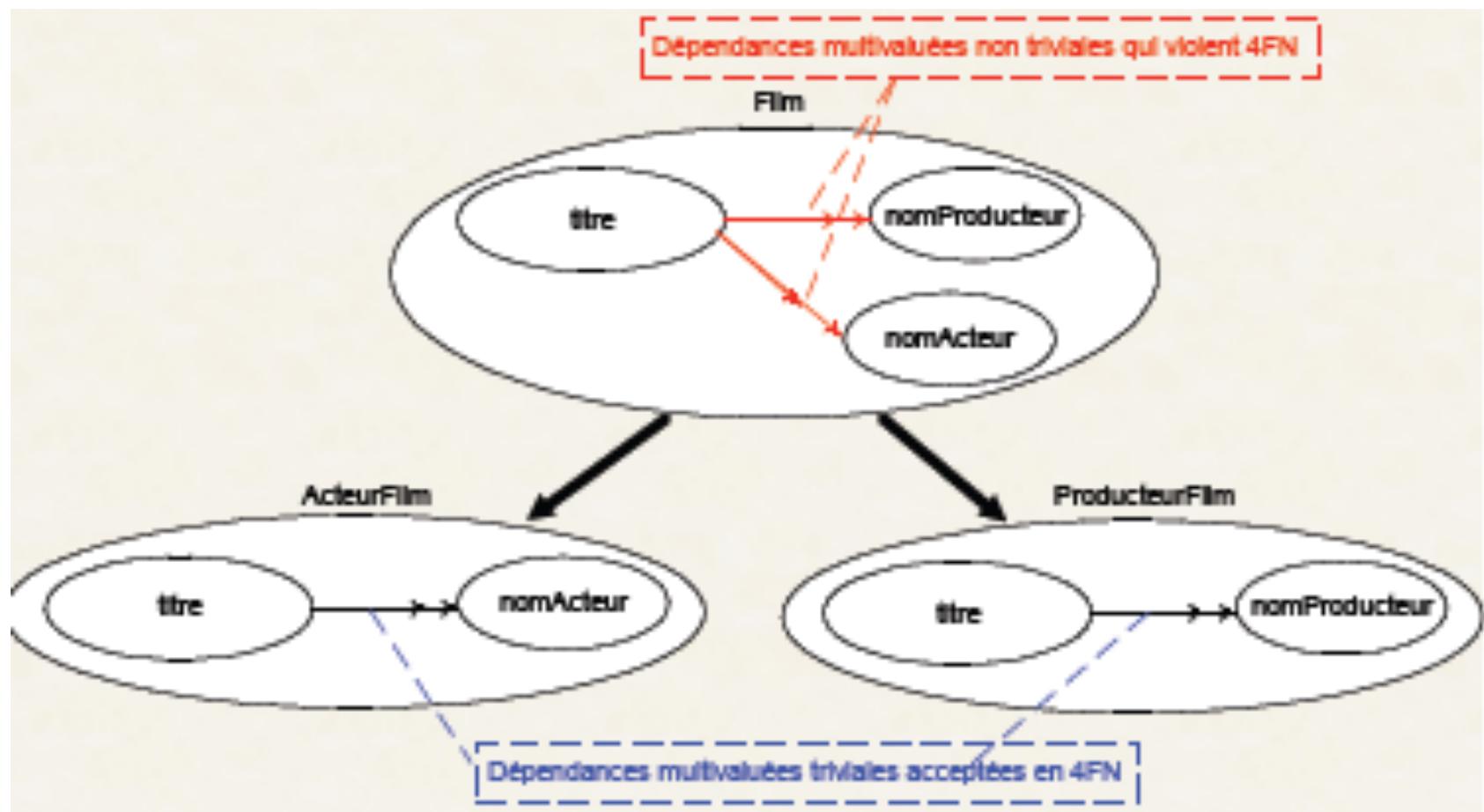
Généralisation directe du patron de décomposition



Exemple : Film



Exemple : Film



Exemple (suite)

Table Film		
titre	nomProducteur	nomActeur
La vie est belle	Elda Ferri	Roberto Benigni
La vie est belle	Elda Ferri	Nicoletta Braschi
La vie est belle	Elda Ferri	Giorgio Cantarini
La vie est belle	Gianluigi Braschi	Roberto Benigni
La vie est belle	Gianluigi Braschi	Nicoletta Braschi
La vie est belle	Gianluigi Braschi	Giorgio Cantarini
Patch Adams	Barry Kemp	Robin Williams
Patch Adams	Barry Kemp	Monica Potter
Patch Adams	Michael Farrell	Robin Williams
Patch Adams	Michael Farrell	Monica Potter
Patch Adams	Marvin Minoff	Robin Williams
Patch Adams	Marvin Minoff	Monica Potter



Table ActeurFilm	
titre	nomActeur
La vie est belle	Roberto Benigni
La vie est belle	Giorgio Cantarini
La vie est belle	Nicoletta Braschi
Patch Adams	Robin Williams
Patch Adams	Monica Potter

Table ProducteurFilm	
titre	nomProducteur
La vie est belle	Elda Ferri
La vie est belle	Gianluigi Braschi
Patch Adams	Barry Kemp
Patch Adams	Michael Farrell
Patch Adams	Marvin Minoff

Dépendance multivaluée triviale

- $A_1, A_2, \dots, A_n \rightarrow\rightarrow_T B_1, B_2, \dots, B_m$ est triviale si :
 - ◆ $\text{Schéma}(T) = \{A_1, A_2, \dots, A_n\} \cup \{B_1, B_2, \dots, B_m\}$
 - ou
 - ◆ $\{B_1, B_2, \dots, B_m\} \subseteq \{A_1, A_2, \dots, A_n\}$
- $A \rightarrow\rightarrow_T B$ est triviale si $A \cup B = T$ ou $B \subseteq A$

Quatrième forme normale (4FN)

- Pour toute dépendance multivaluée **non triviale**, $X \rightarrow\rightarrow^T Y$, le déterminant X est une **superclé** de T
- **Proposition**
 - $4FN \Rightarrow FNBC$

Algorithme de décomposition en 4FN

Procédure *Décomposition4FN* (T, S)

Entrée:

T : une table avec ses dépendances multivaluées

Sortie

S : un schéma relationnel pour T en FNBC

DÉBUT

Ajouter T à S

TANT QU'il y a une table T dans S qui n'est pas en 4FN

Décomposer T selon une dépendance non triviale $X \rightarrow\!\!\!\rightarrow^T Y$

qui ne respecte pas la 4FN

et remplacer T par $T_1(U-Y)$ et $T_2(X \cup Y)$ dans S

FIN TANT QUE

FIN

Dépendances multivaluées emboîtées

- $\text{sigleCours} \rightarrow\!\!\! \rightarrow idEtudiant$ ne tient pas dans $Cours$ mais « surgit » dans $Cours2$!

Table <i>Cours</i>			
sigleCours	idEtudiant	codeProf	nbCours
INF5180	1	p1	2
INF5180	2	p1	3
INF5180	1	p2	4
INF5180	2	p2	1
...			

$idEtudiant, codeProf \rightarrow nbCours$

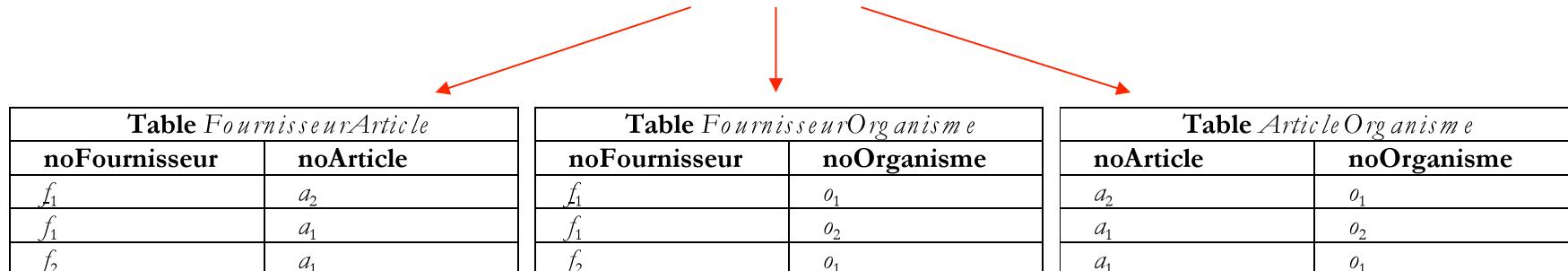
Table <i>NombreCours</i>		
idEtudiant	codeProf	nbCours
1	p1	2
2	p1	3
1	p2	4
2	p2	1

Table <i>Cours2</i>		
sigleCours	idEtudiant	codeProf
INF5180	1	p1
INF5180	2	p1
INF5180	1	p2
INF5180	2	p2
...		

Cinquième forme normale

- Décomposition n -aire ($n > 2$)

Table <i>Fournisseur</i>		
noFournisseur	noArticle	noOrganisme
f_1	a_2	o_1
f_1	a_1	o_2
f_2	a_1	o_1
f_1	a_1	o_1



Dépendance de jointure

- $| \{X_1, X_2, \dots, X_n\}$ dans T si
 - $T = \pi_{x_1}(T) | \pi_{x_2}(T) | \dots | \pi_{x_n}(T)$
- Dépendance de jointure triviale
 - si une des parties, X_i , est l'ensemble de toutes les colonnes de T
- Ex: $| \{(noFournisseur, noArticle), (noFournisseur, noOrganisme), (noArticle, noOrganisme)\}$

$|$ est le signe de jointure naturelle[⊗]

Dépendance de jointure

- **Proposition (DMV cas particulier d 'une DJ)**
 - $X \rightarrow\rightarrow^T Y \Rightarrow | \{(X \cup Y), (X, U-X-Y)\}$
- Exemple de la table *Film*
 - $\text{titre} \rightarrow\rightarrow^T \text{nomProducteur} \Rightarrow$
 $\{\text{(titre, nomProducteur)}, \text{(titre, nomActeur)}\}$
 $= ;\text{titre, nomProducteur}(T) \mid ;\text{titre, nomActeur}(T)$
 $= \text{ProducteurFilm} \mid \text{ActeurFilm}$
 $= T$

Cinquième forme normale (5FN)

- Pour toute dépendance de jointure non triviale $| \{X_1, X_2, \dots, X_n\} \text{ dans } T, \text{ chacun des } X_i$ est une superclé de T
- **Proposition**
 - $5\text{FN} \Rightarrow 4\text{FN}$
- En général :
 - $5\text{FN} \Rightarrow 4\text{FN} \Rightarrow \text{FNBC} \Rightarrow 3\text{FN} \Rightarrow 2\text{FN} \Rightarrow 1\text{FN}$
 - $\text{DF} \Rightarrow \text{DMV} \Rightarrow \text{DJ}$

Dépendance généralisée

- Sous forme de tableau :

noFournisseur	noArticle	noOrganisme
f_1	a_2	o_1
f_1	a_1	o_2
f_2	a_1	o_1
f_1	a_1	o_1

- les premières lignes représentent la règle
 - la dernière ligne représente la conclusion
- les 3er lignes présentes \Rightarrow la 4ième doit y être aussi

Formale normale domaine-clé (FNDC)

- Toutes les contraintes déductibles de
 - domaines
 - clés candidates
- Considère aussi fragmentation horizontale

Remarques au sujet de la normalisation

- Normalisé : pas toujours satisfaisant
- Autres considérations de conception
 - valeurs nulles
 - colonnes dérivées
 - choix des colonnes
 - codage
 - performance en lecture
 - dénormalisation

Étude de cas