

Réseaux bayésiens dynamiques formalismes, caractéristiques, exemples

Sommaire

7.1	Introduction	88
7.2	Modèles graphiques orientés	89
7.3	Réseaux bayésiens dynamiques	90
7.4	Conclusion	93

Résumé

Ce chapitre propose une présentation théorique des réseaux bayésiens dynamiques. Après une brève description des modèles espace-état, quelques rappels théoriques sur les modèles graphiques sont introduits. Ils permettent de définir les réseaux bayésiens dynamiques et de mettre en évidence leur aptitude à modéliser un processus temporel comme la compréhension de la parole.

7.1 Introduction

Les modèles stochastiques que nous avons conçus pour la génération de fragments de structures sémantiques sont à base de réseaux bayésiens dynamiques (DBN). Les bases théoriques définissant les DBN sont rappelées dans ce chapitre.

Comme la plupart des situations liées à l'activité humaine, le processus de compréhension ne peut être étudié sans considérer sa dimension temporelle. La compréhension d'un message impose l'utilisation d'*observations* multiples au cours de sa formulation. L'analyse de ces observations séquentielles a pour but de capturer le sens du segment de message observé en utilisant les connaissances issues des précédents segments. Le sens et les variables sémantiques utilisées pour le représenter n'étant pas observés, la modélisation du processus par des modèles espace-état est adéquate.

Les modèles espace-état intègrent la distinction entre les variables observées (par exemple les mots...) et les variables cachées (par exemple les frames sémantiques associées au fragment de message). Ils permettent de représenter les relations de dépendances et de causalités dans un cadre temporel (Dean et Kanazawa, 1989).

Les variables du modèle étant fixées, un modèle espace-état est complètement défini par la donnée de la distribution de probabilités :

- de l'état initial du système étudié,
- des transitions entre états,
- des observations sur les états cachés.

Les modèles graphiques sont des outils de représentation efficaces des modèles espace-état. Alliant théorie des probabilités et théorie des graphes, les modèles graphiques permettent de représenter de façon factorisée des distributions jointes de probabilités sur un ensemble de variables aléatoires. Une étude exhaustive des capacités de modélisation, d'inférence et d'apprentissage de ces modèles est proposée dans (Jordan, 1998).

Un des atouts des modèles graphiques est leur modularité : un modèle complexe peut être construit en associant des modèles simples. Les graphes sont des objets de représentation intuitifs des variables, de leurs dépendances et des dépendances entre états. La consistance du modèle est assurée par la théorie des probabilités.

Dans le cadre de ce travail, les modèles graphiques étudiés sont des modèles dirigés acycliques, nommés *Réseaux bayésiens* (Pearl, 1986, 1998). Ces réseaux utilisent conjointement connaissances d'experts et observations dans le cadre inférentiel bayésien. Le graphe dirigé acyclique modélisant la structure du réseau est déterminé par l'expert qui fixe les relations de dépendance entre les variables du modèle. La distribution de probabilités jointe sur l'ensemble des variables du modèle est ensuite définie en utilisant les propriétés de décomposition propres aux modèles graphiques rappelées dans la section 7.2 suivante.

7.2 Modèles graphiques orientés

Les notions rappelées dans ce paragraphe sont développées dans de nombreux travaux de référence tels (Pearl, 1998; Charniak, 1991).

Soit X_1, \dots, X_n , variables aléatoires discrètes définies par leur loi jointe P . Ces variables sont représentées par les nœuds $v_i \in V$ du graphe orienté $G(V, E)$ associé au réseau bayésien. La figure 7.1 présente un exemple de graphe de réseau bayésien.

Les arcs $e_i \in E$ du graphe associé au réseau bayésien représentent les dépendances entre les variables. On dit que $u \in V$ est un parent de $v \in V$ si $(u, v) \in E$. L'ensemble des nœuds parents d'un nœud v est noté $pa(v)$. Les fils de v sont les nœuds dont v est un parent. Les descendants de v sont les nœuds fils de fils et leurs descendants.

La structure graphique d'un réseau bayésien satisfait le critère de *d-séparation* : toute variable est indépendante de tout sous-ensemble de ses non-descendants, conditionnellement à ses parents.

En toute généralité, la formule des probabilités composées permet d'écrire :

$$P(X_1, \dots, X_n) = P(X_n | X_{n-1}, X_{n-2}, \dots, X_2, X_1) \dots P(X_2 | X_1) P(X_1) \quad (7.1)$$

$$= P(X_1) \prod_{i=2}^n P(X_i | X_{i-1}, \dots, X_1) \quad (7.2)$$

$pa(X_i)$ étant l'ensemble des variables qui conditionnent X_i , la d-séparation permet d'écrire :

$$P(X_i | X_{i-1}, \dots, X_1) = P(X_i | pa(X_i)) \quad (7.3)$$

D'après (7.2) et (7.3) la loi jointe s'écrit alors sous la forme :

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i)) \quad (7.4)$$

qui définit complètement le réseau bayésien.

La probabilité d'une variable X_i n'est plus calculée conditionnellement à toutes les réalisations de ses prédécesseurs X_{i-1}, \dots, X_1 mais seulement à celles de ses parents, éléments de $pa(X_i)$.

Si le graphe du réseau est complet, soit au pire cas, la factorisation (7.4) est équivalente à la factorisation générale (7.2) à l'ordre des variables près. Dans tous les autres cas, d'autant plus favorables que la densité du graphe est faible, le calcul de la loi jointe est très simplifiée par l'utilisation de la factorisation (7.4).

Considérant l'exemple de réseau bayésien illustré par la figure 7.1, on a $pa(X_1) = pa(X_2) = \emptyset$, $pa(X_3) = \{X_1, X_2\}$, $pa(X_4) = pa(X_5) = \{X_3\}$ et $pa(X_6) = \{X_4, X_5\}$.

La loi jointe définissant le réseau permet donc d'écrire :

$$P(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_1)P(X_2)P(X_3|X_1, X_2)P(X_4|X_3)P(X_5|X_3)P(X_6|X_4, X_5)$$

grâce à la factorisation (7.4).

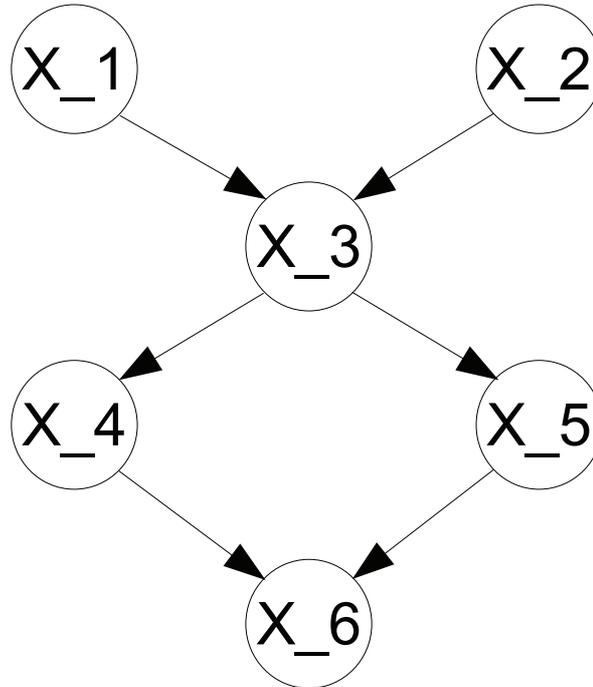


FIGURE 7.1 – Graphe d'un réseau bayésien

Les réseaux bayésiens décrits dans ce paragraphe ne peuvent appréhender que des ensembles finis de variables aléatoires. Ils ne sont donc pas aptes à modéliser le processus de la compréhension orale dans sa dimension temporelle et séquentielle. De nouveaux types de réseaux ont été proposés pour représenter les processus temporels en conservant les atouts des réseaux bayésiens. Il s'agit des *réseaux bayésiens dynamiques* étudiés dans le paragraphe suivant.

7.3 Réseaux bayésiens dynamiques

Les réseaux bayésiens dynamiques (*Dynamic Bayesian Network*, DBN) étendent la notion de réseaux bayésiens à la modélisation de distributions de probabilités sur des ensembles dénombrables de vecteurs aléatoires. Les phénomènes étant modélisés dans un *espace d'états*, ces vecteurs aléatoires représentent les variables observées, cachées et inférées du modèle. Une étude exhaustive des DBN est présentée dans (Mihajlovic et Petkovic, 2001) et (Murphy, 2002).

Seuls les processus stochastiques à temps discret sont étudiés dans ce travail. Le pas

de temps t est incrémenté à chaque nouvelle observation, le modèle proposant ainsi la représentation d'une succession d'événements discrets.

L'aspect *dynamique* d'un DBN est lié à sa faculté de modéliser l'évolution temporelle d'un phénomène. La structure du réseau bayésien choisi pour représenter le phénomène n'évolue pas avec le temps. En revanche, les distributions de probabilités associées aux états successifs de ce réseau sont dépendantes de la succession temporelle des observations. Ces dépendances sont supportées par les variables aléatoires de *transition* entre deux états successifs. Les DBN permettent la modélisation des dépendances entre les variables du réseau bayésien au temps t ainsi que celles liant deux étapes temporelles successives $t - 1$ et t . Un DBN modélise donc un système dynamique.

Le schéma 7.2 décrit un exemple de modèle DBN sur trois étapes temporelles successives. Les arcs pleins reliant les nœuds matérialisent les dépendances entre les variables du réseau à chaque étape de temps tandis que les arcs pointillés indiquent les dépendances entre deux étapes de temps successives.

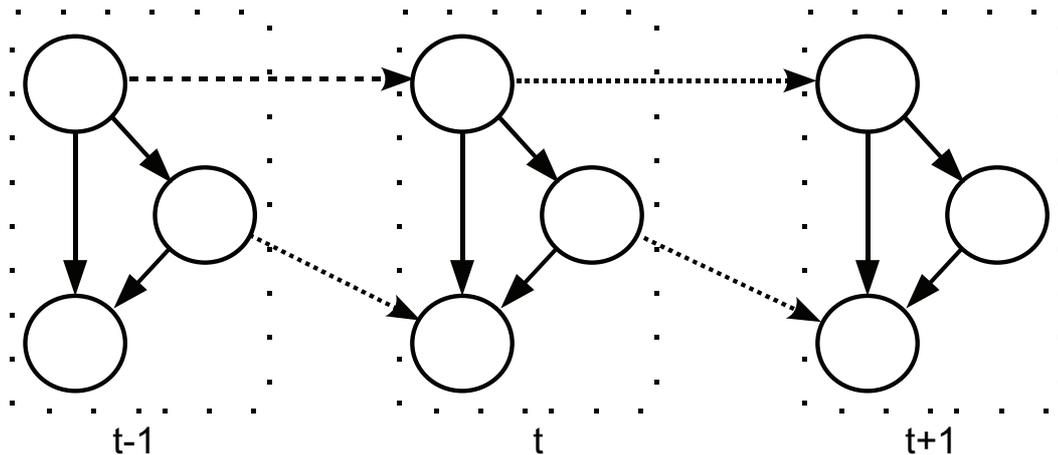


FIGURE 7.2 – Schéma représentant l'évolution d'un processus sur trois étapes temporelles successives $t - 1$, t et $t + 1$.

V_t étant le vecteur aléatoire représentant l'état du système au temps t , un DBN est défini par l'état initial du système et le réseau bayésien temporel décrivant les relations entre deux états successifs.

Formellement, un DBN est donc complètement défini par le couple (B_1, B_{\rightarrow}) (notation partiellement empruntée à (Murphy, 2002)) tel que :

- B_1 est le réseau bayésien définissant la distribution de probabilités initiale $P(V_1)$;
- B_{\rightarrow} est le réseau bayésien associé à deux pas de temps qui définit la distribution de probabilités $P(V_t|V_{t-1})$.

Représentée par un graphe orienté acyclique, $P(V_t|V_{t-1})$ s'écrit sous la forme :

$$P(V_t|V_{t-1}) = \prod_{i=1}^N P(V_t^i | pa(V_t^i))$$

où V_t^i est la i ème composante aléatoire de V_t (associée au nœud i du graphe) et $pa(V_t^i)$ représente les composantes aléatoires dont dépend V_t^i (i.e. les nœuds parents du nœud i dans le graphe de B_{\rightarrow}).

Chaque nœud i est associé à une distribution de probabilités conditionnelle au temps t , $P(V_t^i | pa(V_t^i))$ pour tout $t > 1$.

Dans ce travail, les parents $pa(V_t^i)$ du nœud i au temps t peuvent être des nœuds du même état (temps t) ou des nœuds de l'état précédent (temps $t - 1$). Tout système représenté par un DBN est donc markovien du premier ordre : l'état du système au temps t dépend uniquement de son état au temps $t - 1$. Cette propriété traduit l'indépendance du futur conditionnellement au passé. Ce choix est motivé par le souhait de limiter la complexité des modèles utilisés. En effet, du point de vue théorique, rien n'interdit à un DBN de modéliser des dépendances entre des étapes de temps non successives.

Selon le principe de causalité, les arcs reliant les états successifs sont orientés dans l'ordre temporel. Les modèles considérés dans ce travail sont homogènes dans le temps, i.e. les distributions de probabilités conditionnelles sont invariantes au cours du temps.

La distribution de probabilités jointe associée à un DBN, définie par le déroulement du réseau bayésien B_{\rightarrow} sur T séquences temporelles, s'écrit donc :

$$P(V_{1,\dots,T}) = \prod_{t=1}^T \prod_{i=1}^N P(V_t^i | pa(V_t^i))$$

L'exemple le plus simple de DBN est le HMM pour lequel un état comporte une seule variable cachée et une seule variable observée à chaque étape temporelle. Un HMM est représenté graphiquement sur trois périodes temporelles par la figure 7.3.

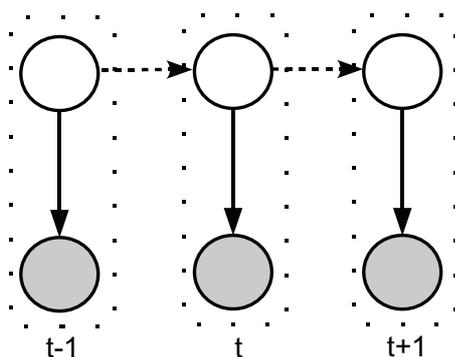


FIGURE 7.3 – Exemple de modèle HMM représenté graphiquement sur trois périodes temporelles

Par convention, les nœuds symbolisant des variables observées sont grisés tandis que les nœuds clairs désignent les variables cachées.

7.4 Conclusion

Ce chapitre a présenté les fondements théoriques des modèles graphiques orientés et en particulier l'aspect dynamique des DBN. Leurs caractéristiques mettent en évidence l'aptitude de ces modèles à modéliser des systèmes dynamiques complexes, tant dans leur dimension temporelle que situationnelle. Ils offrent un cadre théorique et pratique complet pour représenter une large gamme de probabilités conditionnelles mettant en jeu des variables stochastiques discrètes ou continues et réaliser des opérations complexes comme l'inférence sur celles-ci.

Ainsi, la capacité des DBN à modéliser la tâche de compréhension de messages oraux a motivé leur utilisation dans ce travail. Le chapitre suivant [8](#) détaille les raisons de ce choix. Il présente les modèles conçus pour générer les fragments sémantiques sur lesquels repose la compréhension et les paramètres associés à ces modèles.

Chapitre 8

Des réseaux bayésiens dynamiques pour la génération de fragments sémantiques

Sommaire

8.1	Introduction	96
8.2	Modèle compact	98
8.3	Modèle factorisé	101
8.4	Modèle à deux niveaux	105
8.5	Définition et dérivation des fragments sémantiques	108
8.6	Conclusion	110

Résumé

Ce chapitre présente l'approche à base de réseaux bayésiens dynamiques utilisée dans ce travail pour la génération de fragments de structures sémantiques. L'organisation structurelle des modèles utilisés est détaillée. Les paramètres stochastiques choisis sont précisés et les méthodes dédiées à l'apprentissage à partir des données de ces paramètres sont explicitées.

8.1 Introduction

Les réseaux bayésiens dynamiques ont été présentés dans le chapitre précédent. Ce sont des modèles d'une grande flexibilité permettant de représenter des systèmes stochastiques complexes. Leur adaptabilité autorise des modélisations variées qu'il est peu coûteux de faire évoluer. Les DBN, utilisés dans de nombreuses tâches de modélisation de données séquentielles, obtiennent des résultats au niveau de l'état de l'art (voir (Lefèvre, 2007) pour un exemple d'application des DBN au problème de l'interprétation littérale de la parole).

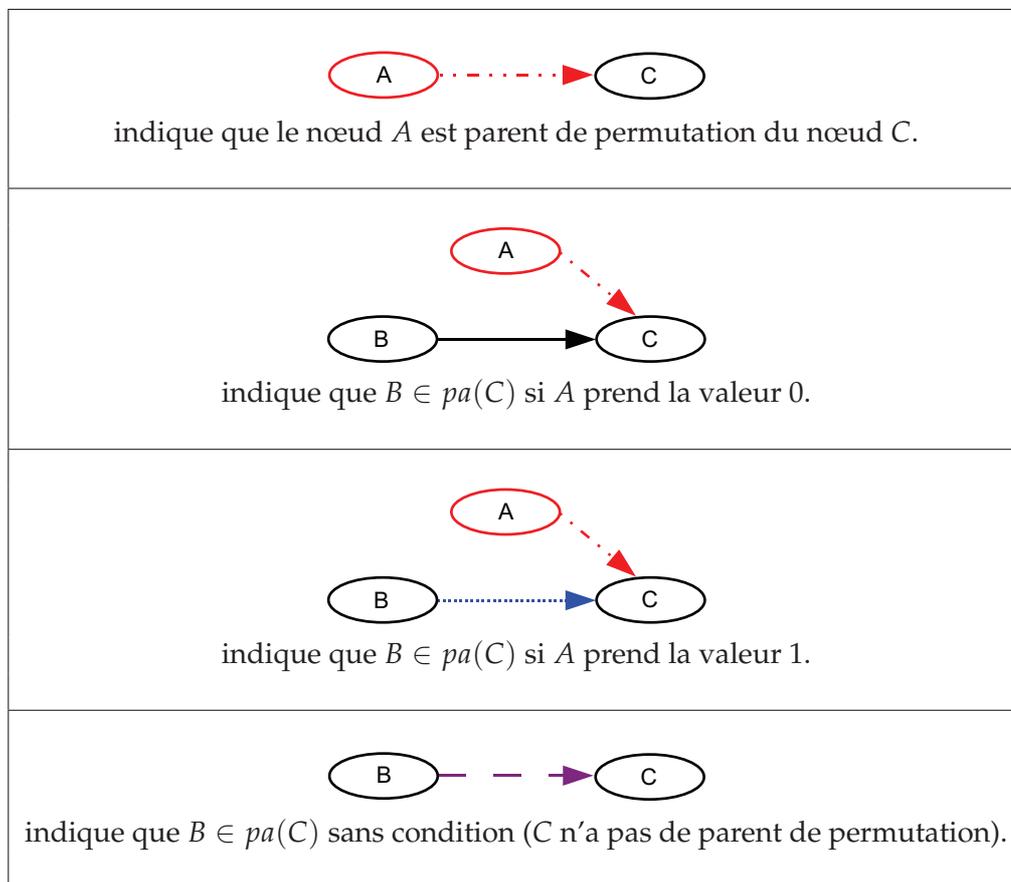
Le contexte dans lequel ce travail a été réalisé est sous-tendu par les théories de la causalité sous incertitude et des réseaux de croyance (Dean et Kanazawa, 1988, 1989). La modélisation du processus de compréhension de la parole pour laquelle nous avons opté repose sur les capacités des DBN à générer les fragments sémantiques associés aux observations des mots fournies par le module de reconnaissance de la parole et aux concepts de base fournies par le module de compréhension littérale.

Le corpus MEDIA (voir CHAP.4) étant manuellement transcrit et annoté en concepts de base, les paramètres des modèles évalués peuvent être appris directement sur les observations. Les modèles mettent en œuvre les variables aléatoires représentant les mots, les concepts de base et les objets sémantiques (frames et FE) ainsi que les transitions entre ces entités. Les objets sémantiques sont portés par des variables cachées tandis que les mots et concepts de base sont considérés comme des observations. Les relations de dépendance entre variables sont établies dans l'esprit d'une influence directe des objets sémantiques sur les mots et concepts de base observés.

Les paragraphes suivants présentent les structures et les paramètres des modèles génératifs à base de DBN utilisés pour la composition des frames dans notre système de compréhension. Dans les graphes proposés, les nœuds symbolisent les variables et les arcs matérialisent les dépendances conditionnelles. Par souci de lisibilité, certains nœuds et certains arcs ne sont pas représentés dans les graphes simplifiés des figures 8.1, 8.3 et 8.5. Les versions complètes des deux premiers modèles sont données dans les figures 8.2 et 8.4.

Dans les graphes simplifiés, seuls deux événements temporels (soit deux mots) sont indiqués. En pratique, ce schéma est répété autant que nécessaire au long de la séquence de mots étudiée, ce qui est indiqué par la flèche circulaire des graphes généraux. Les nœuds sont grisés lorsque les variables sont observées tandis qu'ils restent clairs pour les variables cachées. Les arcs pleins illustrent les dépendances conditionnelles entre les variables. Les arcs pointillés qui indiquent que les dépendances conditionnelles changent en fonction de la valeur d'un *parent de permutation* représenté par un nœud également en pointillés. L'utilisation de parents de permutation dans les réseaux bayésiens (nommés alors *multi-réseaux*) permet de représenter efficacement le changement de conditionnement en fonction de la valeur d'un nœud (Geiger et Heckerman, 1996; Bilmes, 2000). Les valeurs des parents de permutation sont binaires. Elles indiquent dans nos modèles l'état particulier d'un nœud en fonction du contexte.

La légende des graphes de représentation des modèles investigués est donnée ci-après :



Légende des graphes de représentation des modèles DBN étudiés

Toutes les variables sont observées pendant l'entraînement du modèle. Ainsi, les tables de probabilités conditionnelles associées aux arcs sont directement obtenues à partir des observations. Aucune itération EM n'est donc nécessaire.

Le calcul des probabilités de ces tables est réalisé grâce à des modèles de langage factorisés (*Factored Language Models*, FLM) utilisant des techniques de repli parallèle généralisé (*Generalized Parallel Backoff*, GPB) (Bilmes et Kirchhoff, 2003; Kirchhoff et al., 2008). Les FLM sont une extension des modèles de langage classiques dans laquelle les prédictions sont basées sur un ensemble de caractéristiques et non plus seulement sur les précédentes occurrences de la variable. Le repli parallèle généralisé permet d'étendre les procédures de repli standard au cas où des éléments de différents types sont considérées, sans contrainte temporelle imposée : contrairement aux modèles de langage classiques, dans un FLM les éléments intervenant au moment de la prédiction peuvent être pris en compte.

La section suivante 8.2 présente un modèle compact dans lequel les fragments de frames et FE sont représentés par une seule variable composite. Le modèle factorisé qui considère des frames et FE représentés par deux variables distinctes mais simultanément décodées est décrit en 8.3. Enfin, un modèle à deux niveaux dans lequel les frames sont décodées en premier lieu, puis utilisées comme des valeurs observées lors du décodage des FE, est détaillé dans la section 8.4.

8.2 Modèle compact

La figure 8.1 décrit le modèle génératif à base de DBN dans lequel les fragments de frames et FE sont représentés par une seule variable composite **frM**. Le graphe complet du DBN est donné dans la figure 8.2.

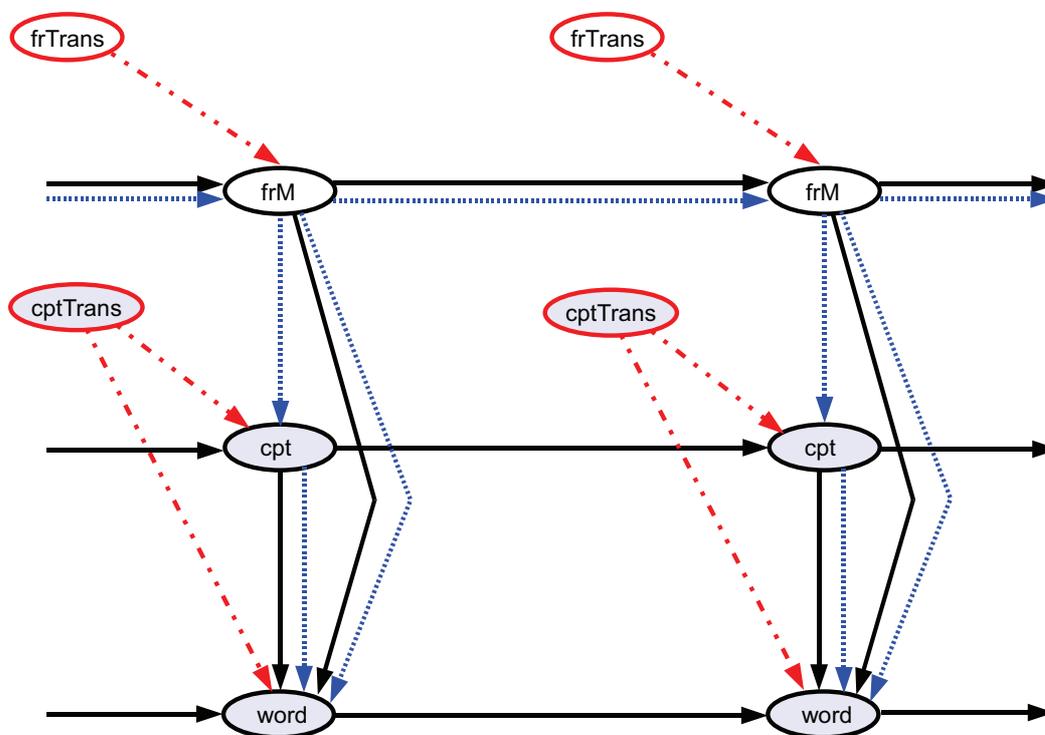


FIGURE 8.1 – frame et FE considérés comme une seule variable non-observée.

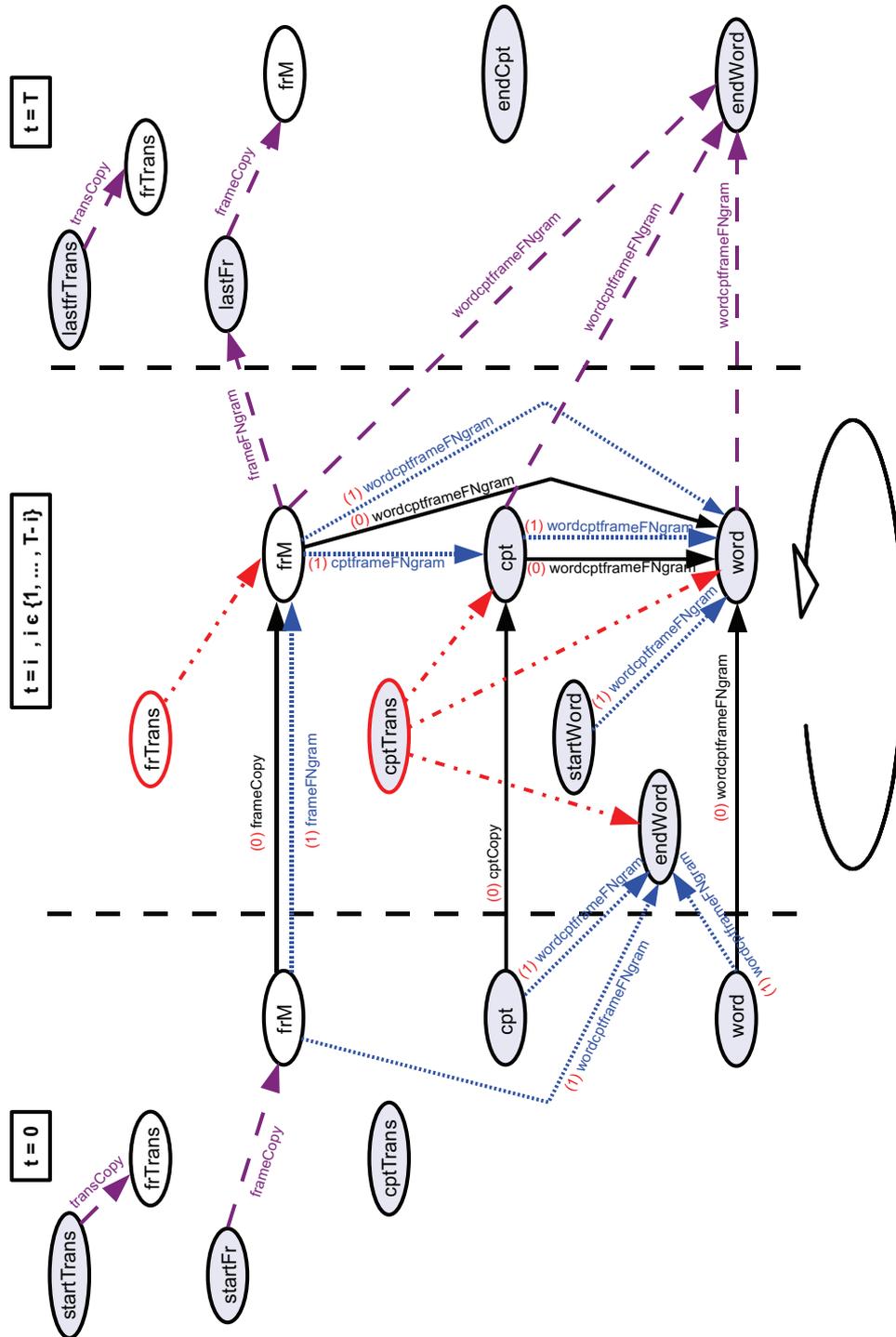


FIGURE 8.2 – Modèle compact *frame/FE*

Pour ce modèle, les variables observées sont :

- les mots : nœuds **word**,
- les concepts de base : nœuds **cpt**,
- les transition entre concepts : nœuds **cptTrans**,
- les variables “outils” :
 - nœuds **startTrans**, **startFr**, **lastfrTrans**, **lastFr** et **endCpt** qui définissent le DBN é l’origine du message (temps $t = 0$) et é la fin du message (temps $t = T$),
 - nœuds **startWord** et **endWord** qui indiquent début et fin d’une séquence de mots associée é un concept.

Les variables cachées sont :

- les fragments de frames-FE : nœuds **frM**,
- les transitions entre ces fragments : nœuds **frTrans**.

La variable transitionnelle cachée représentée par le nœud clair **frTrans** est parent de permutation du nœud **frM**. La permutation porte dans ce cas sur la distribution de probabilités utilisée pour prédire la valeur du nœud **frM** au temps t sachant l’état du modèle é t_1 . Ainsi, si **frTrans** a une valeur nulle, le fragment frame/FE associée au nœud **frM** est identique é celui de son prédécesseur. La table de probabilités conditionnelles utilisée pour déterminer la valeur du fragment frame/FE é l’étape t est la matrice identité de taille $n \times n$ oé n est le cardinal de l’ensemble des fragments rencontrés dans les données d’apprentissage. Cette table est nommée *frameCopy* et mentionnée dans le graphe 8.2.

Lorsque **frTrans** est égal é un, la nouvelle valeur du fragment frame/FE est déterminée en fonction de la probabilité $P(f|f_{-1})$ du fragment f connaissant le fragment précédent f_{-1} . La table de probabilités conditionnelles utilisée est nommée *frameFNgram* dans le graphe 8.2.

La variable transitionnelle observée représentée par le nœud grisé **cptTrans** est parent de permutation des nœuds **cpt** et **word**. Cette variable prend la valeur 0 si le concept ne change pas de l’étape $t - 1$ é t , et la valeur 1 dans le cas contraire.

En l’absence de changement de concept observé, le concept associé au nœud **cpt** courant est identique é celui de son prédécesseur selon la table de probabilités *cptCopy*, identité de taille $m \times m$ oé m est le cardinal de l’ensemble des concepts rencontrés dans les données d’apprentissage.

Le mot associé au nœud **word**, observé, détermine les valeurs de probabilités associées é son conditionnement par le mot précédent, le concept courant et le fragment de frame/FE courant selon la table de probabilités conditionnelles *wordcptframeFNgram* mentionnée en 8.2

Lors du changement de concept observé, **cptTrans** prend la valeur 1 et le concept associé au nœud **cpt** courant détermine les valeurs de probabilités associées au conditionnement du concept par le fragment frame/FE. La table de probabilités conditionnelles utilisée dans ce cas, nommée *cptframeFNgram*, est mentionnée dans le graphe 8.2.

Le mot associé au nœud **word**, observé, est conditionné selon *wordcptframeFNgram* par le concept et le fragment de frame/FE courants et par le nœud **startWord** qui in-

dique l'entrée dans une nouvelle séquence de mots associée au concept courant.

Les tables de probabilités conditionnelles correspondant aux arcs du graphe 8.1 sont produites par les implémentations des FLM données ci-après. Les variables FFE , C et W représentent respectivement un fragment frame/FE, un concept et un mot tandis que h fixe la longueur de l'historique ($h = -1$ pour un bigramme).

- distribution conditionnelle sur les séquences de fragments de F/FE ;

$$P(FFE) \simeq \prod P(ffe|ffe_h) : \text{frameFNgram}$$

- distribution conditionnelle sur les séquences de concepts conditionnées par les fragments frame/FE. Le GPB est effectué dans l'ordre $\{c_h, ffe\}$;

$$P(C|FFE) \simeq \prod P(c|c_h, ffe) : \text{cptframeFNgram}$$

- distribution conditionnelle sur les séquences de mots conditionnées par les concepts et les fragments frame/FE. Le GPB est effectué dans l'ordre $\{w_h, c, ffe\}$.

$$P(W|C, FFE) \simeq \prod P(w|w_h, c, ffe) : \text{wordcptframeFNgram}$$

Le choix de représenter les fragments de frames et FE par une seule variable est essentiellement motivé par la réduction de la complexité du décodage. La cardinalité de l'ensemble des valeurs possibles de frame/FE est limitée au nombre de fragments de frame/FE observés dans les données d'apprentissage. Cependant, cette approche conduit à l'utilisation de liens déterministes entre frames et FE.

8.3 Modèle factorisé

La figure 8.3 décrit le modèle génératif à base de DBN dans lequel les fragments de frames et FE sont représentés par deux variables distinctes **frM** et **feM**, simultanément décodées. Le graphe complet du DBN est donné dans la figure 8.4.

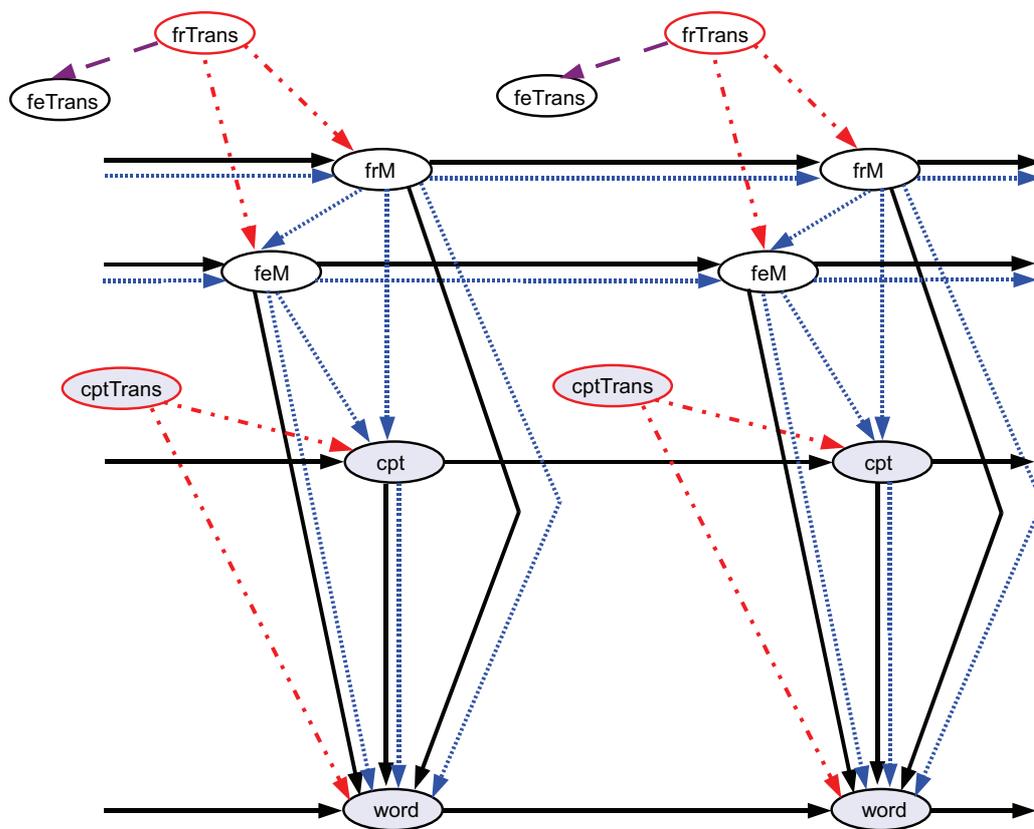


FIGURE 8.3 – *frame et FE considérés comme deux variables non-observées.*

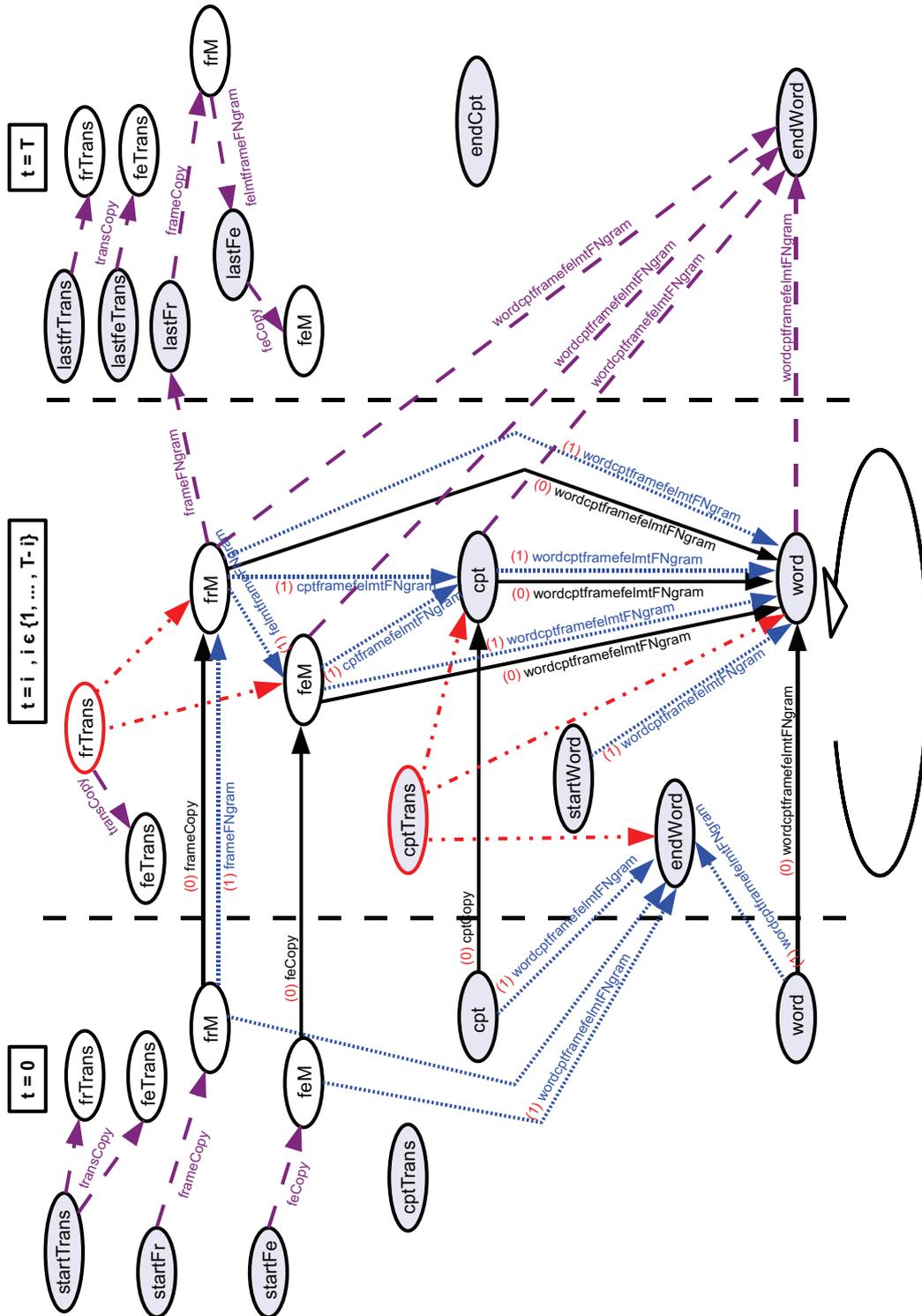


FIGURE 8.4 – Modèle factorisé

Pour ce modèle, les variables observées sont :

- les mots : nœuds **word**,
- les concepts de base : nœuds **cpt**,
- les transition entre concepts : nœuds **cptTrans**,
- les variables “outils” :
 - nœuds **startTrans**, **startFr**, **startFe**, **lastfrTrans**, **lastfeTrans**, **lastFr**, **lastFe** et **endCpt** qui définissent le DBN é l’origine du message (temps $t = 0$) et é la fin du message (temps $t = T$),
 - nœuds **startWord** et **endWord** qui indiquent début et fin d’une séquence de mots associée é un concept.

Les variables cachées sont :

- les fragments de frames : nœuds **frM**,
- les fragments de FE : nœuds **feM**,
- les transitions entre ces fragments : nœuds **frTrans** et **feTrans**.

Variationnelles transitionnelles et parents de permutation se comportent comme décrit en 8.2.

Les tables de probabilités conditionnelles correspondant aux arcs du graphe 8.3 sont produites par les implémentations des FLM données ci-après. Les variables F , FE , C et W représentent respectivement un fragment frame, un fragment FE, un concept et un mot tandis que h fixe la longueur de l’historique ($h = -1$ pour un bigramme).

- distribution conditionnelle sur les séquences de fragments de frames ;

$$P(F) \simeq \prod P(f|f_h) : \text{frameFNgram}$$

- distribution conditionnelle sur les séquences de fragments de FE conditionnées par les fragments de frames. Le GPB est effectué dans l’ordre $\{f_{e_h}, f\}$;

$$P(FE|F) \simeq \prod P(fe|f_{e_h}, f) : \text{felmtframeFNgram}$$

- distribution conditionnelle sur les séquences de concepts conditionnées par les fragments de frames et les fragments de FE. Le GPB est effectué dans l’ordre $\{c_h, fe, f\}$;

$$P(C|FE, F) \simeq \prod P(c|c_h, fe, f) : \text{cptframefelmtFNgram}$$

- distribution conditionnelle sur les séquences de mots conditionnées par les concepts, les fragments de frames et les fragments de FE. Le GPB est effectué dans l’ordre $\{w_h, c, fe, f\}$.

$$P(W|C, FE, F) \simeq \prod P(w|w_h, c, fe, f) : \text{wordcptframefelmtFNgram}$$

Le modèle factorisé permet de considérer les ambiguïtés des liens entre frames et FE en leur attribuant des probabilités et testant chaque combinaison au cours du décodage. Les combinaisons non rencontrées dans les données d'apprentissage sont évaluées grâce é des techniques de repli. Dans nos modèles le GPB utilise la technique de discount absolu (ou Kneser-Ney) (Kneser et Ney, 1995). évidemment, cette approche est coûteuse et la complexité du modèle est telle qu'il est nécessaire d'utiliser un algorithme sous-optimal de recherche en largeur (de type recherche en faisceau, *beam search*) pendant le décodage.

8.4 Modèle é deux niveaux

Le modèle é deux niveaux dans lequel les frames sont décodées en premier lieu, puis utilisées comme des valeurs observées lors du décodage des FE est illustré par la figure 8.5. Le graphe développé du premier niveau de ce modèle est semblable é celui de la figure 8.2, où la variable **frM** représente les fragments de frames. Le graphe développé du second niveau est semblable é celui de la figure 8.4 où les variables **frM** et **frTrans** sont observées.

Pour le premier niveau :

Les variables observées sont :

- les mots : nœuds **word**,
- les concepts de base : nœuds **cpt**,
- les transition entre concepts : nœuds **cptTrans**,

Les variables cachées sont :

- les fragments de frames : nœuds **frM**,
- les transitions entre ces fragments : nœuds **frTrans**.

Pour le second niveau :

Les variables observées sont :

- les mots : nœuds **word**,
- les concepts de base : nœuds **cpt**,
- les transition entre concepts : nœuds **cptTrans**,
- les fragments de frames : nœuds **frM**,
- les transitions entre ces fragments : nœuds **frTrans**

Les variables cachées sont :

- les fragments de FE : nœuds **feM**

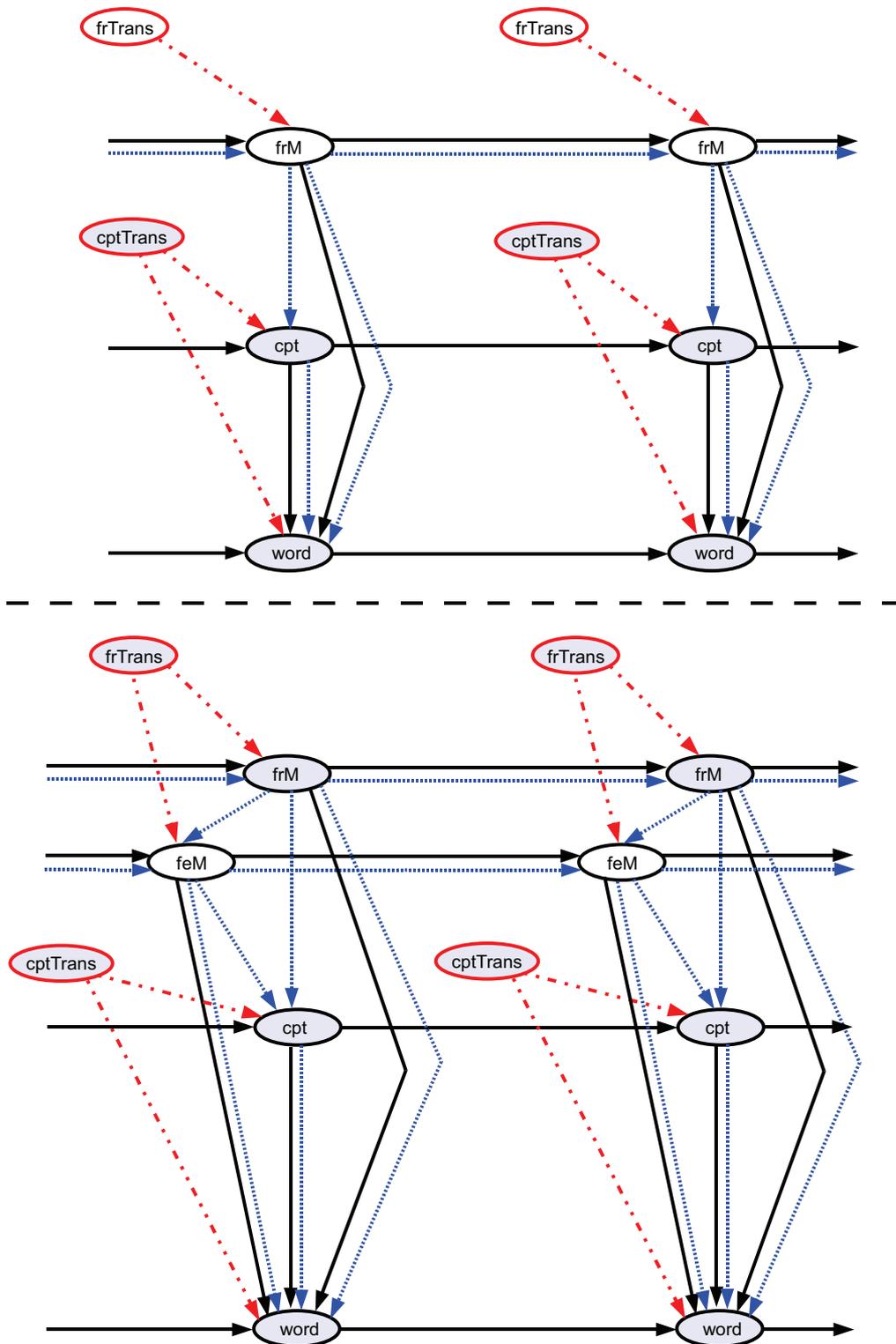


FIGURE 8.5 – Décodage à deux niveaux des frames et FE

Variationnelles et parents de permutation se comportent comme décrit en 8.2.

Les tables de probabilités conditionnelles correspondant aux arcs du graphe 8.5 sont produites par les implémentations des FLM données ci-après. Les variables F , FE , C et W représentent respectivement un fragment frame, un fragment FE, un concept et un mot tandis que h fixe la longueur de l'historique ($h = -1$ pour un bigramme).

★ *Premier niveau :*

- distribution conditionnelle sur les séquences de fragments de frames ;

$$P(F) \simeq \prod P(f|f_h) : \text{frameFNgram}$$

- distribution conditionnelle sur les séquences de concepts conditionnées par les fragments de frames. Le GPB est effectué dans l'ordre $\{c_h, f\}$;

$$P(C|F) \simeq \prod P(c|c_h, f) : \text{cptframeFNgram}$$

- distribution conditionnelle sur les séquences de mots conditionnées par les concepts et les fragments de frames. Le GPB est effectué dans l'ordre $\{w_h, c, f\}$.

$$P(W|C, F) \simeq \prod P(w|w_h, c, f) : \text{wordcptframeFNgram}$$

★ *Second niveau :*

- distribution conditionnelle sur les séquences de fragments de frames observés ;

$$P(\hat{F}) \simeq \prod P(\hat{f}|\hat{f}_h) : \text{frameFNgram}$$

- distribution conditionnelle sur les séquences de fragments de FE conditionnées par les fragments de frames observés. Le GPB est effectué dans l'ordre $\{fe_h, \hat{f}\}$;

$$P(FE|\hat{F}) \simeq \prod P(fe|fe_h, \hat{f}) : \text{felmtframeFNgram}$$

- distribution conditionnelle sur les séquences de concepts conditionnées par les fragments de frames observés et les fragments de FE. Le GPB est effectué dans l'ordre $\{c_h, \hat{f}, fe\}$;

$$P(C|\hat{F}, FE) \simeq \prod P(c|c_h, \hat{f}, fe), \text{ GPB dans l'ordre } \{c_h, \hat{f}, fe\} : \text{cptframefelmtFNgram}$$

- distribution conditionnelle sur les séquences de mots conditionnées par les concepts, les fragments de frames observés et les fragments de FE. Le GPB est effectué dans l'ordre $\{w_h, c, \hat{f}, fe\}$.

$$P(W|C, \hat{F}, FE) \simeq \prod P(w|w_h, c, \hat{f}, fe) : \text{wordcptframefelmtFNgram}$$

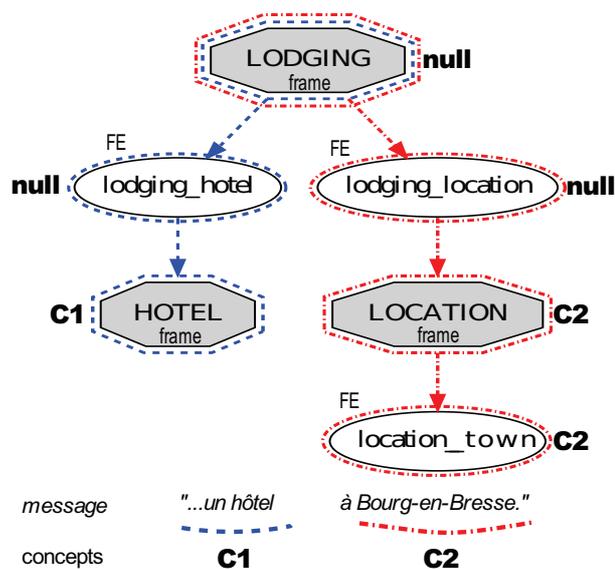


FIGURE 8.6 – Branches projetées associées à la séquence “... un hôtel à Bourg-en-Bresse”.

où le chapeau indique les variables ayant des valeurs fixées.

Bien qu'également sous-optimale, l'approche à deux niveaux de décodage propose un modèle de complexité moindre que celle du modèle factorisé tout en conservant l'aptitude à décoder des frames et FE reliés par des liens non déterministes.

8.5 Définition et dérivation des fragments sémantiques

La représentation des frames et FE étant hiérarchique, des situations de recouvrement peuvent se produire lors de la détermination des frames et des FE associés à un concept. Cela arrive principalement lorsque plusieurs frames ou FE ont été déclenchés par le même concept mais également lorsque les processus d'inférence et de composition ont créé des structures de frames et FE imbriquées reliées au même concept. Pour résoudre ce problème, un algorithme de projection d'arbre est appliqué. Il est décrit ci-dessous et son pseudo-code est donné dans l'algorithme 1.

La projection est réalisée sur l'annotation en frames et FE, structurée en arbre, de la phrase complète. Elle permet de définir des sous-branches de l'arbre associées à un seul concept. Partant d'une feuille de l'arbre, une branche de frame/FE est obtenue en agrégeant les valeurs de nœuds pères (frame ou FE) aussi longtemps qu'ils sont associés au même concept (ou à aucun). Les arêtes des branches sont définies par les liens d'appartenance des FE aux frames et par les liens de typage des FE en frame (un FE prenant une frame pour valeur).

Par exemple, la séquence de mots "... un hôtel é Bourg-en-Bresse", proposée dans la figure 8.6, entraîne la création des branches projetées

HOTEL-lodging_hotel-LODGING et

location_town-LOCATION-lodging_location-LODGING.

Les branches sont exploitées selon les différents modèles DBN. Dans le cas du modèle compact, les branches sont directement considérées comme des classes composées. Dans les modèles factorisés et é deux niveaux, frames et FE sont séparés pour produire deux ensembles de classes distinctes.

t] Algorithme de projection d'arbres

Entrée : $\{c_i\}$ séquence de concepts associés é la phrase, \mathcal{T} arbre de frames et FE représentant le message

Sortie : \mathcal{B} ensemble des branches

```

1:  $\mathcal{B} \leftarrow \emptyset$ 
2: pour tout  $c \in \{c_i\}$  faire
3:   branche  $b_c \leftarrow \emptyset$ 
   Génération des branches principales
4:   pour tout  $l \in \text{feuilles}(\mathcal{T}, c)$  faire
5:      $b_c.\text{ajouter}(\text{extraire\_branche}(l))$ 
6:   fin pour
   Contrôle des branches internes
7:   pour tout  $n \in \text{noeuds}(\mathcal{T}, c)$  faire
8:     si  $n \notin b_c$  alors
9:        $b_c.\text{ajouter}(\text{extraire\_branche}(n))$ 
10:    fin si
11:   fin pour
12:    $\mathcal{B} \leftarrow b_c$ 
13: fin pour
14: retourner  $\mathcal{B}$ 

```

Fonction `extraire_branche`

Entrée : c_i concept, n noeud

Sortie : branche $b \leftarrow \emptyset$

```

15: répéter
16:    $b. = n$ 
17:    $n \leftarrow n.\text{père}()$ 
18: jusqu'à  $!(n \text{ ET } (n.\text{concept} \in \{c_i, \text{null}\}))$ 
19: retourner  $b$ 

```

end

8.6 Conclusion

Les trois modèles DBN dédiés à la génération de fragments sémantiques proposés dans ce travail se distinguent par leurs structures.

Le modèle compact considère des variables représentant des fragments sémantiques composés de frames et de FE. Son atout principal est de réduire la complexité générale du décodage. Le choix d'une variable unique pour la représentation conjointe des frames et FE est cependant une contrainte forte puisqu'elle induit l'utilisation de liens déterministes entre frames et FE.

A l'opposé, le modèle factorisé est celui qui offre la plus grande liberté de combinaison au sein des fragments sémantiques. En effet, chaque association frames - FE est évaluée au cours du décodage. Ce modèle est en contrepartie celui de complexité maximale.

Le modèle à deux niveaux présente l'avantage d'une approche non déterministe des liens frames - FE tout en ayant une complexité inférieure à celle du modèle factorisé.

L'apprentissage des distributions de probabilité utilisées par ces modèles s'appuie sur un algorithme de décomposition des arbres sémantiques en sous-branches conceptuelles.

Les modèles détaillés dans ce chapitre sont évalués sur le corpus MEDIA. Les conditions d'évaluation et les résultats obtenus sont présentés dans le chapitre suivant 9.