

La structure morpho-syntaxique des phrases dans le résumé automatique

Compression de phrases narratives

Mehdi Yousfi-Monod, Violaine Prince

LIRMM, UMR 5506

161 rue Ada

34392 Montpellier Cedex 5 - France



9 juin 2005

Plan

2

- 1 La compression de phrases
- 2 L'analyseur morpho-syntaxique SYGFRAN
- 3 Notre approche : étude de la fonction et de la position des constituants
- 4 L'architecture de notre système
- 5 L'expérimentation
- 6 Conclusion, perspectives

La compression de phrases

La compression de phrases consiste, de manière générale, à supprimer certains mots ou constituants des phrases souvent un peu trop longues, en conservant, du mieux que possible, les éléments les plus pertinents.

La compression de phrases

La compression de phrases consiste, de manière générale, à supprimer certains mots ou constituants des phrases souvent un peu trop longues, en conservant, du mieux que possible, les éléments les plus pertinents.

La compression de phrases

La compression de phrases consiste, de manière générale, à supprimer certains mots ou constituants des phrases souvent un peu trop longues, en conservant, du mieux que possible, les éléments les plus pertinents.

La compression de phrases

La compression de phrases consiste, de manière générale, à supprimer certains mots ou constituants des phrases souvent un peu trop longues, en conservant, du mieux que possible, les éléments les plus pertinents.

La compression de phrases

La compression de phrases consiste, de manière générale, à supprimer certains mots ou constituants des phrases souvent un peu trop longues, en conservant, du mieux que possible, les éléments les plus pertinents.

La compression de phrases

La compression de phrases consiste, de manière générale, à supprimer certains mots ou constituants des phrases souvent un peu trop longues, en conservant, du mieux que possible, les éléments les plus pertinents.



La compression de phrases consiste à supprimer certains mots des phrases, en conservant les éléments les plus pertinents.

G. Grefenstette 1998

- importance des constituants basée principalement sur leur nature
- suppression des moins importants
- application : simplification de texte destinés à être lus par des mal voyant

G. Grefenstette 1998

- importance des constituants basée principalement sur leur nature
- suppression des moins importants
- application : simplification de texte destinés à être lus par des mal voyant

→ cohérence grammaticale faible

→ peu d'informations syntaxiques sont utilisées

- ressources utilisées :

- ① corpus de couples <phrase, phrase compressée>
- ② lexique (Jing, McKeown 98)
- ③ Wordnet (Miller et al. 90)
- ④ parseur ESG (McCord 90)

- ressources utilisées :

- ① corpus de couples <phrase, phrase compressée>
- ② lexique (Jing, McKeown 98)
- ③ Wordnet (Miller et al. 90)
- ④ parseur ESG (McCord 90)

- importance des constituants basée sur :

- ① leur rôle thématique
- ② le type du verbe utilisé comme prédicat du constituant
- ③ la connexité syntaxique des constituants avec leur contexte
- ④ la probabilité qu'un humain puisse les supprimer

- ressources utilisées :

- ① corpus de couples <phrase, phrase compressée>
- ② lexique (Jing, McKeown 98)
- ③ Wordnet (Miller et al. 90)
- ④ parseur ESG (McCord 90)

- importance des constituants basée sur :

- ① leur rôle thématique
- ② le type du verbe utilisé comme prédicat du constituant
- ③ la connexité syntaxique des constituants avec leur contexte
- ④ la probabilité qu'un humain puisse les supprimer

→ choix de contraction en accord avec ceux des humains (à 81 %)

→ contraction : 67 % (humains : 58 %)

K. Knight and D. Marcu 2002

- apprentissage sur un corpus de couples
<document, résumé>
- passage des phrases avec le parseur Collins
- modèles utilisés :
 - ① probabilistique (canal bruité)
 - ② basé sur la décision

K. Knight and D. Marcu 2002

- apprentissage sur un corpus de couples
<document, résumé>
- passage des phrases avec le parseur Collins
- modèles utilisés :
 - ① probabilistique (canal bruité)
 - ② basé sur la décision

→ cohérence grammaticale limitée

→ nature et position des consituants utilisées, mais pas la fonction

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)
- règles de transformations d'éléments structurés, basées sur les règles de la grammaire française

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)
- règles de transformations d'éléments structurés, basées sur les règles de la grammaire française
- transforme une phrase (texte brut) en un arbre syntaxique (élément structuré) enrichi d'informations sur les constituants

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)
- règles de transformations d'éléments structurés, basées sur les règles de la grammaire française
- transforme une phrase (texte brut) en un arbre syntaxique (élément structuré) enrichi d'informations sur les constituants

Avantages

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)
- règles de transformations d'éléments structurés, basées sur les règles de la grammaire française
- transforme une phrase (texte brut) en un arbre syntaxique (élément structuré) enrichi d'informations sur les constituants

Avantages

- 1 rapidité : $O(k * n * \log_2(n))$ (k : nombre de règles, n taille de la donnée textuelle)

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)
- règles de transformations d'éléments structurés, basées sur les règles de la grammaire française
- transforme une phrase (texte brut) en un arbre syntaxique (élément structuré) enrichi d'informations sur les constituants

Avantages

- 1 rapidité : $O(k * n * \log_2(n))$ (k : nombre de règles, n taille de la donnée textuelle)
- 2 robustesse : > 35 % des cas de syntaxe du français correctement analysés, pour les autres cas, analyse partielle mais exploitable

SYGFRAN

7

- analyseur morpho-syntaxique basé sur le système opérationnel SYGMART (Chauché 84)
- règles de transformations d'éléments structurés, basées sur les règles de la grammaire française
- transforme une phrase (texte brut) en un arbre syntaxique (élément structuré) enrichi d'informations sur les constituants

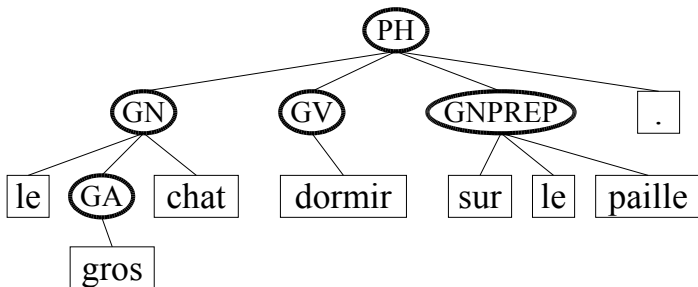
Avantages

- 1 rapidité : $O(k * n * \log_2(n))$ (k : nombre de règles, n taille de la donnée textuelle)
- 2 robustesse : > 35 % des cas de syntaxe du français correctement analysés, pour les autres cas, analyse partielle mais exploitable
- 3 production d'un arbre syntaxique : par opposition à un simple marquage linéaire classique

SYGFRAN : un petit exemple

Le gros chat dort sur la paille.

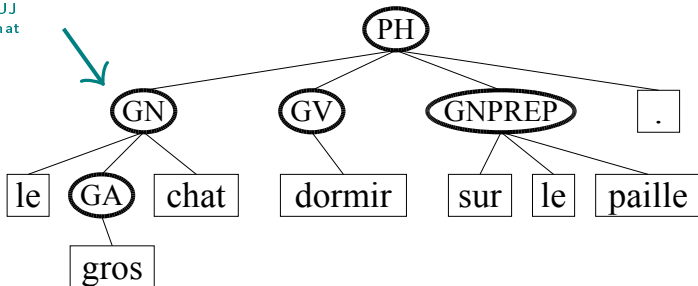
SYGFRAN : un petit exemple



Le gros chat dort sur la paille.

SYGFRAN : un petit exemple

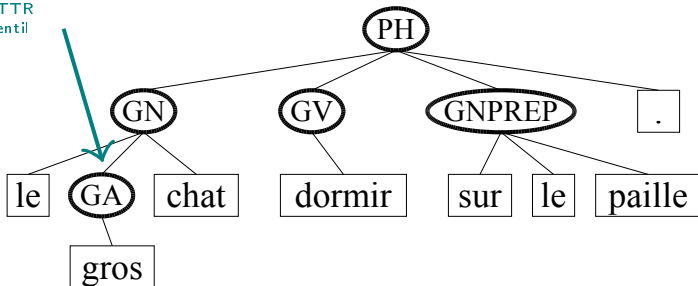
CAT=N
 SOUSN=NCOM
 K=GN
 FS=SUJ
 UL=chat
 ...



Le gros chat dort sur la paille.

SYGFRAN : un petit exemple

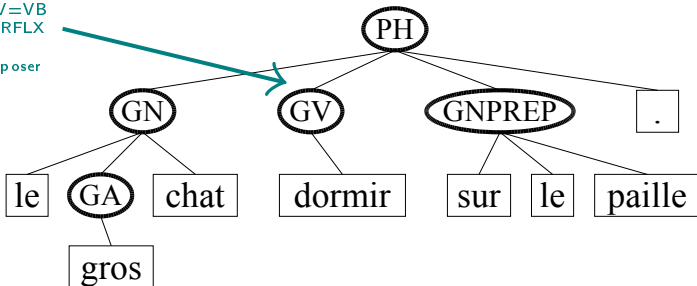
CAT=ADJOINT
 SOUSA=ADNOM
 K=GA
 FS=ATTR
 UL=gentil
 ...xs



Le gros chat dort sur la paille.

SYGFRAN : un petit exemple

NUM=SIN
 PERS=(1,3)
 CAT=V
 SOUSV=VB
 TYP=RFLX
 K=GV
 UL=reposer
 ...



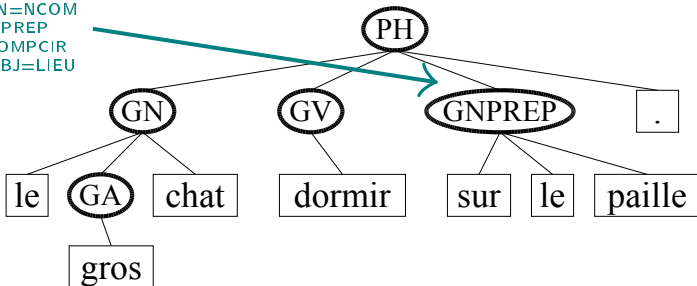
Le gros chat dort sur la paille.

SYGFRAN : exemple

SYGFRAN : un petit exemple

8

GNR=FEM
 NUM=SIN
 CAT=N
 SOUSN=NCOM
 K=GNPREP
 FS=COMPCIR
 SEMOBJ=LIEU
 ...



Le gros chat dort sur la paille.

- fonctions syntaxiques de "gouvernement" (Mel'čuk)

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.
- trois granularités :

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.
- trois granularités :
 - ① phrase :
 - proposition principale > propositions relatives

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.
- trois granularités :
 - ① phrase :
 - proposition principale > propositions relatives
 - ② proposition :
 - sujets et verbes > compléments d'objet > compléments circonstanciels

Constituants gouverneurs

- fonctions syntaxiques de "gouvernement" (Mel'čuk)
 - selon la fonction des constituants, leurs importances grammaticale et sémantique varient
 - Le gros chat dort sur la paille.
- trois granularités :
 - ① phrase :
 - proposition principale > propositions relatives
 - ② proposition :
 - sujets et verbes > compléments d'objet > compléments circonstanciels
 - ③ constituant nominal :
 - noms > complément de noms > adjectifs épithètes

Le genre du texte

L'importance des constituants dépend aussi du genre du texte

L'importance des constituants dépend aussi du genre du texte

- article scientifique → difficile : les éléments d'information ont une grande importance dans l'argumentation scientifique,

L'importance des constituants dépend aussi du genre du texte

- article scientifique → difficile : les éléments d'information ont une grande importance dans l'argumentation scientifique,
- texte narratif (contes, romans, ...) → contraction possible : de nombreuses informations descriptives non indispensables,

L'importance des constituants dépend aussi du genre du texte

- article scientifique → difficile : les éléments d'information ont une grande importance dans l'argumentation scientifique,
- texte narratif (contes, romans, ...) → contraction possible : de nombreuses informations descriptives non indispensables,
- autres genres → en cours d'expérimentation.

Les constituants les moins importants

- 1 les compléments circonstanciels

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- 2 les épithètes

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- 2 les épithètes
 - le cas de l'article défini

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- 2 les épithètes
 - le cas de l'article défini
 - le cas des relatives

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- 2 les épithètes
 - le cas de l'article défini
 - le cas des relatives
- 3 les appositions

Les constituants les moins importants

- 1 les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- 2 les épithètes
 - le cas de l'article défini
 - le cas des relatives
- 3 les appositions
 - bons candidats : groupes nominaux, pronoms, relatives

Les constituants les moins importants

- ① les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- ② les épithètes
 - le cas de l'article défini
 - le cas des relatives
- ③ les appositions
 - bons candidats : groupes nominaux, pronoms, relatives
 - plus difficile : propositions participiales présent et passé

Les constituants les moins importants

- ① les compléments circonstanciels
 - bons candidats : lieu (parfois), manière, cause
 - le cas du verbe être
 - le cas de la phrase interrogative
- ② les épithètes
 - le cas de l'article défini
 - le cas des relatives
- ③ les appositions
 - bons candidats : groupes nominaux, pronoms, relatives
 - plus difficile : propositions participiales présent et passé
 - impossible : proposition infinitive

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important
- compression par passes en fonction du rapport
<taille/perte de contenu> désiré

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important
- compression par passes en fonction du rapport $\langle \text{taille/perte de contenu} \rangle$ désiré

→ Le gros chat dort sur la paille.

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important
- compression par passes en fonction du rapport $\langle \text{taille/perte de contenu} \rangle$ désiré

→ Le **gros** chat dort sur la paille.
épithète

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important
- compression par passes en fonction du rapport
<taille/perte de contenu> désiré

→ Le gros chat dort sur la paille.

Compression par élagage syntaxique

Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important
- compression par passes en fonction du rapport
<taille/perte de contenu> désiré

→ Le gros chat dort **sur la paille**.
CC de lieu

Compression par élagage syntaxique

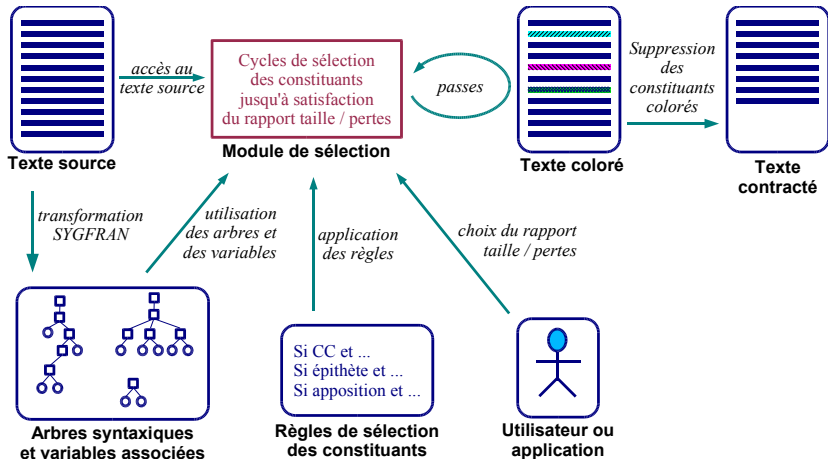
Le principe

- détermination de l'importance selon le genre du texte
→ création de règles de sélection des constituants à supprimer
- suppression des constituants du moins important au plus important
- compression par passes en fonction du rapport $\langle \text{taille/perte de contenu} \rangle$ désiré

→ Le gros chat dort sur la paille.

Architecture de notre système

13



L'expérimentation

- texte de test : conte Haïtien

L'expérimentation

- texte de test : conte Haïtien
- raison du choix : texte narratif correctement analysé (structure de l'arbre) par SYGFRAN

L'expérimentation

- texte de test : conte Haïtien
- raison du choix : texte narratif correctement analysé (structure de l'arbre) par SYGFRAN
- contraction obtenue : 66% du texte d'origine

L'expérimentation

- texte de test : conte Haïtien
- raison du choix : texte narratif correctement analysé (structure de l'arbre) par SYGFRAN
- contraction obtenue : 66% du texte d'origine
- nombre de règles créées : 4

L'expérimentation

- texte de test : conte Haïtien
- raison du choix : texte narratif correctement analysé (structure de l'arbre) par SYGFRAN
- contraction obtenue : 66% du texte d'origine
- nombre de règles créées : 4
- cohérence grammaticale obtenue : parfaite

L'expérimentation

- texte de test : conte Haïtien
- raison du choix : texte narratif correctement analysé (structure de l'arbre) par SYGFRAN
- contraction obtenue : 66% du texte d'origine
- nombre de règles créées : 4
- cohérence grammaticale obtenue : parfaite
- cohérence sémantique obtenue : très correcte

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement
- affinage des règles possible

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement
- affinage des règles possible
- nécessité d'expérimenter sur plus de textes et de genres de textes

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement
- affinage des règles possible
- nécessité d'expérimenter sur plus de textes et de genres de textes
- comparaison difficile avec les autres approches
→ compression de phrases et non pas résumé complet
→ analyse du français (avec les règles actuelles)

Travaux actuels

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement
- affinage des règles possible
- nécessité d'expérimenter sur plus de textes et de genres de textes
- comparaison difficile avec les autres approches
→ compression de phrases et non pas résumé complet
→ analyse du français (avec les règles actuelles)

Travaux actuels

- expérimentations

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement
- affinage des règles possible
- nécessité d'expérimenter sur plus de textes et de genres de textes
- comparaison difficile avec les autres approches
→ compression de phrases et non pas résumé complet
→ analyse du français (avec les règles actuelles)

Travaux actuels

- expérimentations
- informations sur le contexte

Discussion sur les résultats et perspectives

- qualité de la contraction fortement dépendante de SYGFRAN
→ SYGFRAN évolue continuellement
- affinage des règles possible
- nécessité d'expérimenter sur plus de textes et de genres de textes
- comparaison difficile avec les autres approches
→ compression de phrases et non pas résumé complet
→ analyse du français (avec les règles actuelles)

Travaux actuels

- expérimentations
- informations sur le contexte
- résumé inter-phrases : suppression des paraphrases, exemples, explications

Merci beaucoup pour votre attention.
Merci pour votre attention.
Merci.