

# **Etude contrastive en corpus des connecteurs causaux (En-Fr)**

Bruno Cartoni et Sandrine Zufferey

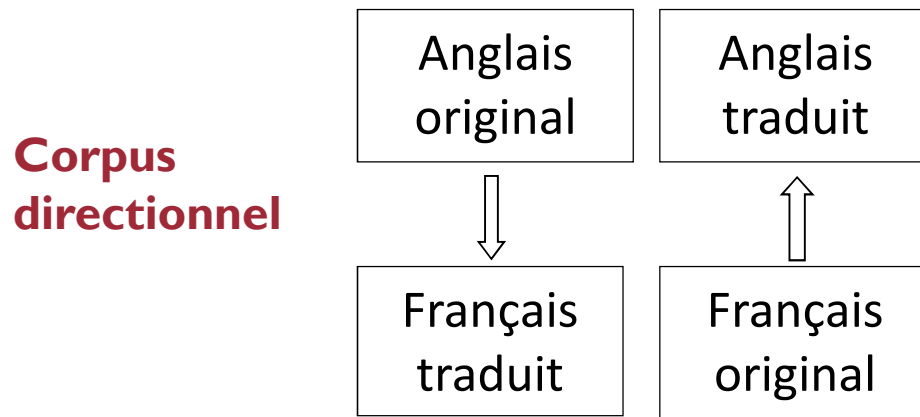
# Motivation

---

- ▶ **Projet COMTIS: Amélioration de la cohérence textuelle en traduction automatique**
- ▶ **Etude empirique multilingue des marques de cohérence**
  - ▶ Temps verbaux, Pronoms, **Connecteurs**
- ▶ **Connecteurs causaux:**
  - ▶ Relation « backward causal»: **Ordre inversé des arguments**  
conséquence **CONNECTEUR** cause  
I fell **BECAUSE** he pushed me
- ▶ **English:** *because, since, for, given that, as*
- ▶ **French:** *parce que, puisque, car, étant donné que, comme*

# Etude empirique basée sur corpus

- ▶ Création de **corpus parallèles directionnels** à partir du corpus multilingue Europarl



- ▶ Etudes contrastive en corpus
  - ▶ Comparaison des deux originaux
  - ▶ Utilisation de la traduction pour éclairer l'original
  - ▶ Comparaison des deux traductions

# Corpus parallèle

<SPEAKER ID=35 LANGUAGE="EN" NAME="Cox">  
Madam President, if the vote records correctly how my Group voted I shall not, and cannot, object to that. If your ruling is that I cannot give an explanation of vote, I accept that but with reservations.

<SPEAKER ID=40 LANGUAGE="EN" NAME="Simpson">  
Madam President, first of all I should like to thank Mr Koch for his report which has, at its heart, the issue of transport safety. The report looks at the issue of harmonising the examination requirements for safety advisors working in the areas of transportation of dangerous goods by road, rail and inland waterway. I congratulate him on his excellent report.

<SPEAKER ID=35 LANGUAGE="EN" NAME="Cox">  
Madame la Présidente, si le procès-verbal reflète correctement le vote de mon groupe, je n'ai et n'aurai aucune objection à formuler. Si votre décision est que je ne puis pas donner d'explication de vote, je l'accepte, mais avec certaines réserves.

<SPEAKER ID=40 LANGUAGE="EN" NAME="Simpson">  
Madame la Présidente, je voudrais avant tout remercier M. Koch de son rapport dans la question de la sécurité des transports occupe une place centrale. Il envisage l'harmonisation du niveau des exigences applicables à l'examen des conseillers à la sécurité pour le transport de marchandises dangereuses par route, par rail ou par voie navigable. Je le félicite de son excellent rapport.

## ► Corpus directionnels En → Fr et Fr → En : 1996-1999

	En → Fr	Fr → En
Statements	5507	4004
Phrases	60'006	45'380
Mots	~ 1,4 millions	~ 1,1 millions

# Connecteur causaux

---

## ▶ Europarl (1996-1999)

- ▶ Statistique dans les parties « langue source »

	Freq.	/100'000 words
<b>parce que</b>	971	82
<b>car</b>	897	76
<b>puisque</b>	534	45
<b>comme (causal)</b>	68	6
<b>étant donné que</b>	26	2

	Freq.	/100'000 words
<b>because</b>	1670	119
<b>since (causal)</b>	135	10
<b>as</b>	144	10
<b>given that</b>	70	5
<b>for</b>	57	4

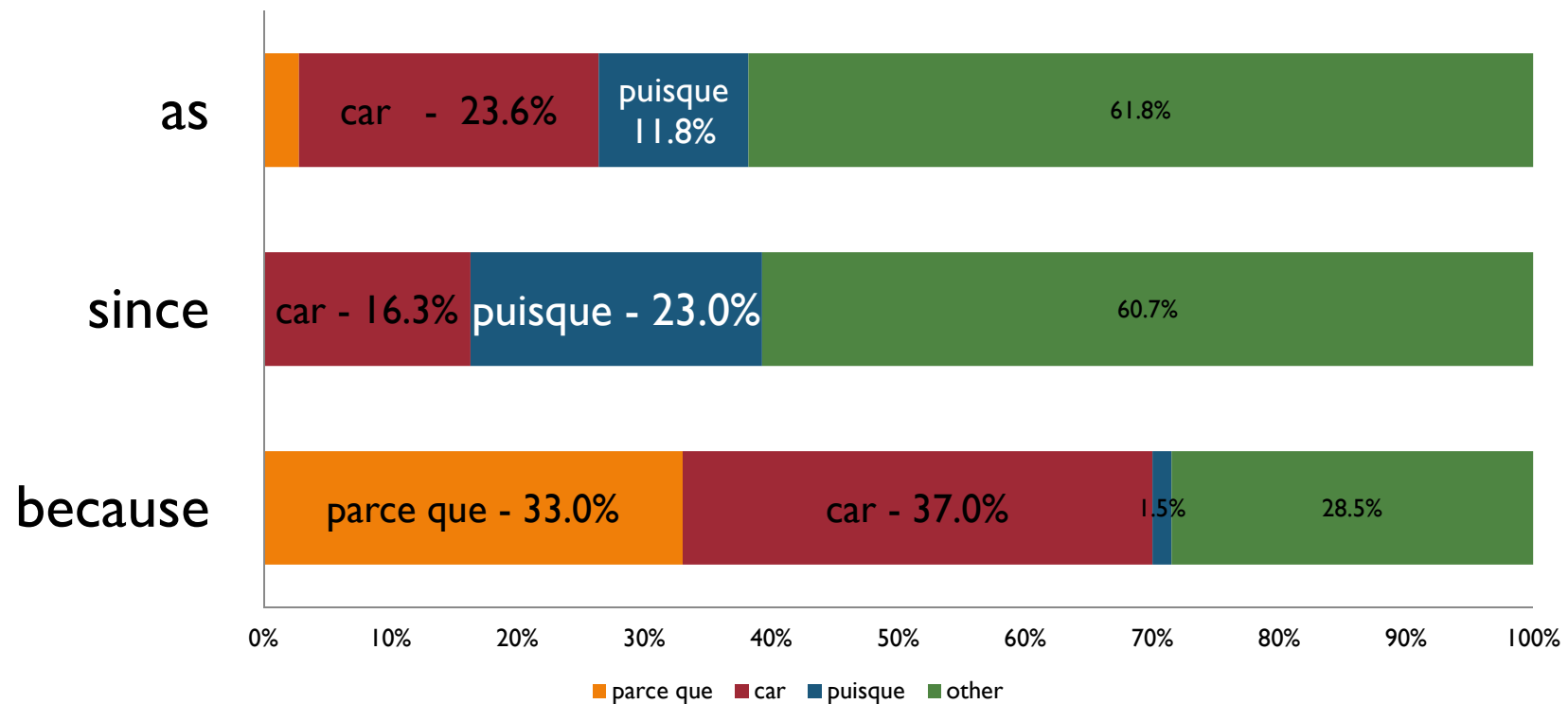
## ▶ Pour notre étude

- ▶ Fréquence > 100
- ▶ Echantillon de 200
- ▶ *parce que, car, puisque*      *because, since, as*

# Translation spotting

---

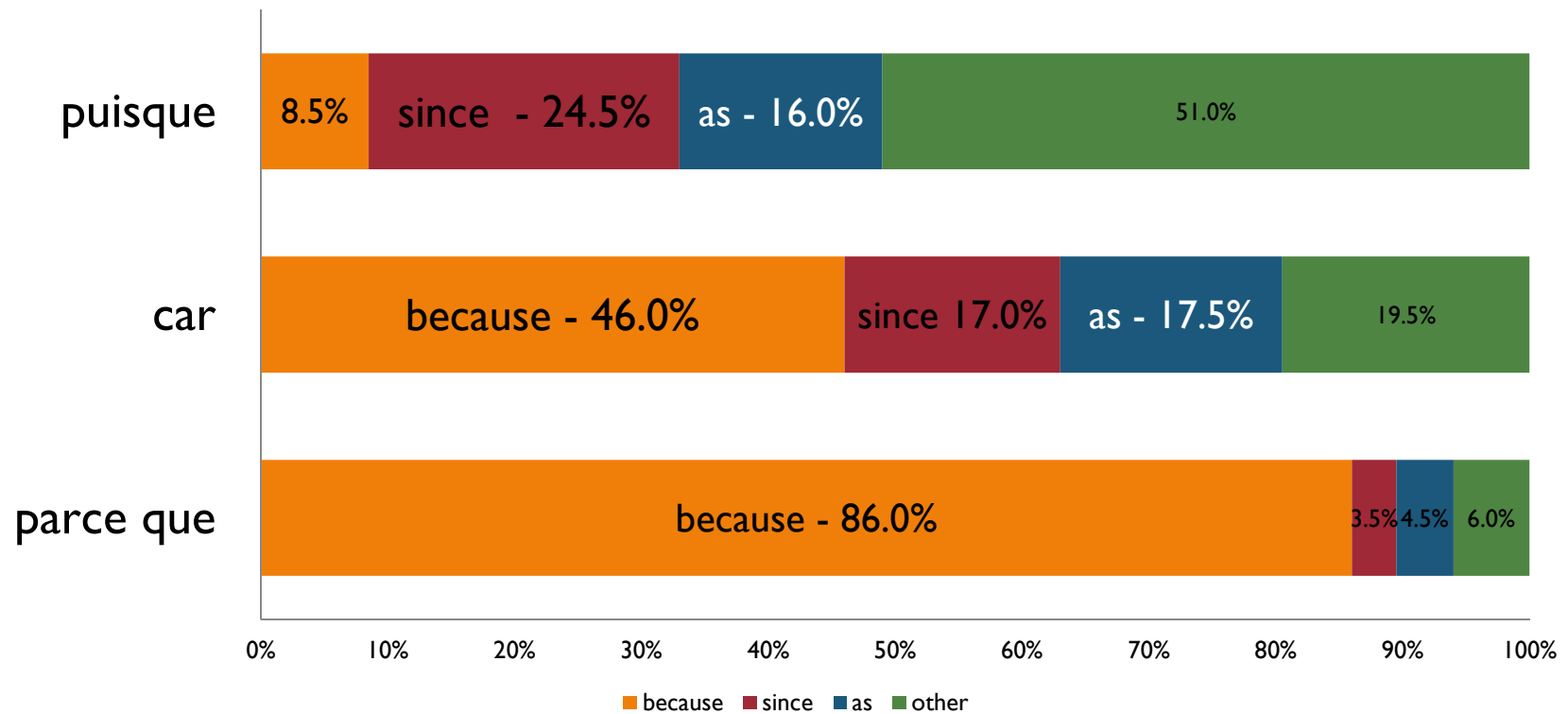
## ► Connecteurs anglais traduits en français



# Translation spotting

---

## ► Connecteurs français traduits en anglais



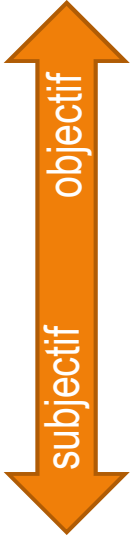
# Conclusion intermédiaire

---

- ▶ Aucune équivalence exacte → divergence de traduction
- ▶ Certains connecteurs ont une large gamme de possibilité de traduction
- ▶ Ces traductions sont-elles interchangeable ?
  - ▶ (Zufferey 2012) : “puisque” et “car” presque jamais interchangeables
- ▶ Y’a-t-il d’autres **traits distinctifs** qui permettent de prédire ces divergences ?
  - ▶ *Domaine d’usage* : objectif vs. subjectif
  - ▶ *Status de l’information* : donné (given) vs. nouveau (new)
  - ▶ *Registre de langue*: oral vs. écrit, formel vs. informel

# Autres traits distinctifs: domaine d'usage

---

- 
- ▶ **Domaine d'usage:**
    - ▶ Relation dans le monde réel : *domaine du contenu*
      - ▶ *La neige fond parce que la température monte.*
    - ▶ Relation entre une croyance et la conclusion dans l'esprit du locuteur: *domaine épistémique*
      - ▶ *Alex est malade, parce qu'il n'est pas venu au travail aujourd'hui.*
    - ▶ Relation impliquant des actes de parole: *domaine des actes de langage*
      - ▶ *Dépêchons-nous, parce qu'il se fait tard.*
  - ▶ La plupart des connecteurs causaux ont un usage *prototypique*
    - ▶ *puisque, car, comme* → subjectif
    - ▶ *parce que* → objectif

# Autres traits distinctifs: status de l'information

---

- ▶ Statut informationnel: nouveau ou donné
  - ▶ donné = mutuellement manifeste (Sperber & Wilson 1986).
- ▶ *Usage prototypique:*
  - ▶ *parce que, car // because, for* → nouveau
  - ▶ *puisque, comme, étant donné que // since, as, given that* → donné
- ▶ Registre et modalité
  - ▶ *for* (archaic), *seeing that* (informal speech), *car* (formel)
  - ▶ *car* (écrit), *comme* (oral)
- ▶ → exclu de notre étude (homogénéité d'Europarl)

# Traits distinctifs en contraste

---

	objectif	Subjectif	new	given
because	+	+	+	
since		+		+
as	?	?		+
parce que	+	+(speech)	+	
car		+	+	
Puisque		+		+

- ▶ Les domaines d'usage et le statut de l'information influencent-ils la traduction ?
- ▶ → double annotation (avec discussion des cas divergents)

# Connecteurs anglais

---

BECAUSE	Total
objectif/nouveau	47
objectif/donné	12
subjectif/nouveau	122
subjectif/donné	19
	<b>200</b>

Objectif : 29.5% } nouveau: 84.5%  
Subjectif : 70.5%

SINCE	Total
objectif/nouveau	9
objectif/donné	2
subjectif/nouveau	69
subjectif/donné	55
	<b>135</b>

Subjectif : 92 %

AS	Total
objectif/nouveau	26
objectif/donné	8
subjectif/nouveau	72
subjectif/donné	35
	<b>141</b>

Subjectif : 75.8 % (nouveau)

# Connecteurs anglais

BECAUSE	Total	car	parce que	Other
objectif/nouveau	47	3	27	17
objectif/donné	12	4	3	5
subjectif/nouveau	122	62	30	30
subjectif/donné	19	5	6	8
	<b>200</b>	<b>74</b>	<b>66</b>	<b>60</b>

SINCE	Total	puisque	étant donné que	car	other
objectif/nouveau	9	1	1	1	6
objectif/donné	2	0	0	0	2
subjectif/nouveau	69	10	10	19	30
subjectif/donné	55	20	17	2	16
	<b>135</b>	<b>31</b>	<b>28</b>	<b>22</b>	<b>54</b>

AS	Total	car	étant donné que	puisque	other
objectif/nouveau	26	4	6	2	14
objectif/donné	8	0	2	2	4
subjectif/nouveau	72	26	7	4	35
subjectif/donné	35	4	6	9	16
	<b>141</b>	<b>34</b>	<b>21</b>	<b>17</b>	<b>69</b>

# Connecteurs français

---

PARCE QUE	Total
objectif/nouveau	78
objectif/donné	15
subjectif/nouveau	91
subjectif/donné	16
	200

Objectif : 46.5%  
 Subjectif : 53.5%  
 nouveau: 84.5%

CAR	Total
objectif/nouveau	16
objectif/donné	6
subjectif/nouveau	155
subjectif/donné	23
	200

Subjectif : 89 %  
 nouveau: 85.5%

PUISQUE	Total
objectif/nouveau	1
objectif/donné	1
subjectif/nouveau	81
subjectif/donné	117
	200

Subjectif : 99 %  
 nouveau: 41 %  
 donné : 59 %

# Connecteurs français

PARCE QUE	Total	because	as	since	other
objectif/nouveau	78	70	3	1	4
objectif/donné	15	14	0	0	1
subjectif/nouveau	91	70	5	6	10
subjectif/donné	16	13	1	0	2
	200	172	9	7	12

CAR	Total	because	as	since	for	other
objectif/nouveau	16	10	2	2	1	1
objectif/donné	6	3	3	0	0	0
subjectif/nouveau	155	73	26	21	20	15
subjectif/donné	23	6	4	11	2	0
	200	92	35	34	23	26

PUISQUE	Total	since	because	as	other
objectif/nouveau	1	1	0	0	0
objectif/donné	1	1	0	0	0
subjectif/nouveau	81	35	15	17	12
subjectif/donné	117	48	17	32	19
	200	87	33	49	31

# Conclusion

---

- ▶ Les données en corpus confirme:
  1. La préférence de certains connecteurs pour un domaine d'usage et un status informationnel particulier.
  2. L'influence des domaines d'usage et du status informationnel sur le choix de traduction

# 1. Préférence des connecteurs

---

	objectif		subjectif	
<b>nouveau</b>	because	<b>23.5%</b>	because	61%
	since	6.5%	since	51%
	as	18.5%	as	51%
	parce que	<b>39%</b>	parce que	45.5%
	car	8%	car	<b>77.5%</b>
	puisque	0.5%	puisque	40.5%
<b>donné</b>	because	6%	because	9.5%
	since	1.5%	since	<b>41%</b>
	as	5.5%	as	25%
	parce que	7.5%	parce que	8%
	car	3%	car	11.5%
	puisque	0.5%	puisque	<b>58.5%</b>

## 2. Conséquence sur la traduction

---

- ▶ < 10% d'usage dans la traduction

	objectif/nouveau	subjectif/nouveau	subjectif/donné
because	parce que	car	
since		car	puisque, étant donné que
as	étant donné que, dans la mesure où	car	puisque, étant donné que
parce que	because	because	
car		because	since
puisque		since	since, as

Merci !