

# CONTRIBUTION OF MOVE STRUCTURE TO AUTOMATIC GENRE IDENTIFICATION

## An Annotated Corpus of French Tourism Websites

---

Rémi Cardon, Trang Pham Tran Hanh, Julien Zakhia Doueïhi & Thomas François  
LREC-COLING 2024

CENTAL – University of Louvain

## GOAL :

Exploring the potential benefit of move structure for automatic genre identification

## HYPOTHESIS :

Informing NLP models with move structure may help in performing automatic genre identification.

## CONTRIBUTIONS :

1. Corpus of tourism websites, in French, annotated with move types
2. Experiments on the potential benefit of move structure in automatic genre identification

1. Move structure
2. Automatic genre identification
3. Resource creation
4. Experiments
5. Conclusion

### GENRE and MOVE STRUCTURE

1. **Discourse Genre** : a class of communicative events : the main criterion that turns a collection of communicative events into a genre is represented by some shared set of communicative purposes (Swales, 1990, p. 45).
2. **Move structure** : Discourse is organized in units (“moves”) fulfilling coherent communicative functions in texts (Swales, 2004, p. 228-229).

### AUTOMATIC GENRE IDENTIFICATION

**Biber's work (1988)** : Several studies have been carried out on automatic genre identification using quantitative and statistical techniques.

**Features analysed** : syntactic features (PoS tags, counting, nominalization), type-token ratios, simple token-level measures (Kuzman and Ljubešić, 2023).

**A gap** : We found only one paper investigating move structure in automatic genre identification (Wu et al., 2006).

## **Tourism Website Homepages**

30 travel agencies (TA) websites

30 Travel blogs (TB)

30 websites of Public Tourism Information (PTI)

30 websites that present Points of Interest (POI)

**Origins** : Belgium, Canada, France and Switzerland.

Same theme but different communication purpose.

## ANNOTATION GUIDELINES

**Steps to do :** Step 1 : Open the website's link on your desktop PC browser, using full screen mode.

Step 2 : Identify each block of text on the home page, following the reading order (left to right, top to bottom).

Step 3 : Identify the move type for each distinct text block.

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 1 : Attracting attention**

- Attract the reader's attention as they enter the homepage (Often in the form of slogans).
- Example : *L'art de voyager ; Trouvez le voyage idéal! ; Mon entreprise, votre avenir.*



## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 2 : Greeting**

- Welcome Internet users.
- Example : *Bienvenue sur le site web de \*Company name\**;  
*Bienvenue!*; *Bienvenue chez \*Company name\**.

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 3 : Identifying sender**

- Identify the web-owner, often achieved through a logo or text containing the web-owner's information.
- Annotate the textual part of the unit if it's a logo.
- Example : *Office de tourisme Vallées de la Braye et de l'Anille ; Coeur de Picard (C) 2020.*

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 4 : Indicating content structure**

- A clear overview of the content of the web site, often referred to as the main menu.
- Annotate all menus displayed on the home page
- Example of menu : *Découvrez-Sortez-Séjournez-Dinoplagne-Infos pratiques-Nos suggestions-Boutique*

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 5 : Detailing (selected) content**

- Offer more detailed information about the topics listed in the main menu.
- A device for news presentation and public image creation (international/national news or news of the self-promotional kind, such as financial results, product news, or latest events where the web-owner is involved)
- A presentation of detailed information about the company/organization/person : professional activities, services offered, etc.
- Example : *Explorer. Une balade en pleine forêt, une datcha romantique, un moment passé au spa, un bon plat gastronomique... Découvrez les activités et idées de sorties en Sartre ! À l'écoute du brame du cerf. Apprendre à fabriquer du savon au lait de jument Montrueil-le-Henri. Plus d'explorations*

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 6 : Establishing credentials**

- Establish a trustworthy image of the web-owner.
- The move can be defined by (1) rational, logical discourse supported by figures, graphs, or percentages (logos), (2) speeches promoting the brand/product/services and its reputation (ethos), and (3) speeches highlighting the advantages for Internet users/consumers of coming to the site or taking advantage of the services offered on the site (pathos).
- Example : *POURQUOI VACANCES SUR MESURE? - Assurance Spéciale Covid-19 - Assistance 7J/7 - Conseils d'experts passionnés - Frais de dossier offerts - Suivi personnalisé - Entreprise à taille humaine*

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 7 : Establishing contact**

- Offer ways for the reader to contact the sender.
- Example : *Réservez votre prochain voyage - Contact ; Contactez-nous ; N'hésitez pas à nous envoyer vos souhaits, afin que nous les transformions en réalité - Contactez-nous.*

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 8 : Establishing a (discourse) community**

- Enables loyal or frequent web users to establish communities revolving around the web site (often realised through a private user space or interaction through external social networks).
- Example : *Suivez-nous sur Facebook ; Recevez notre newsletter ; Abonnez-vous à notre newsletter ; Dernières nouvelles sur Instagram - Suivez-nous.*

## ANNOTATION GUIDELINES

### **Move structure terminology** (Askehave and Nielsen, 2005)

#### **Move 9 : Promoting an external organization**

- Promotes another company, product, etc.
- Example : *Découvrir les beautés de l'Amérique du Nord avec Air Canada!*; *Nos sites partenaires : Les Mondes Polaires, Tirawa, Watch and Cow, Concours du Photographe Voyageur.*



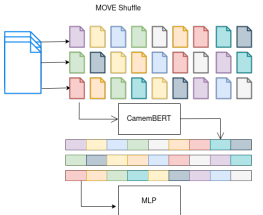
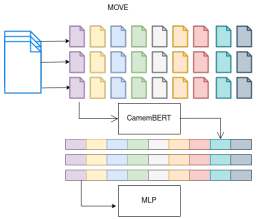
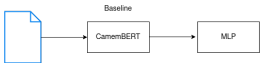
## ANNOTATION PROCESS

- Two annotators
- 2 annotations' test before annotation campaign
- Cohen's  $\kappa = 0.73$  : reliable annotation

Text	TA	TB	PTI	POI
Doc.	419.97 ± 273.18	659.57 ± 478.97	340.61 ± 211.80	277.06 ± 183.20
Move 1	13.86 ± 11.85	11.61 ± 7.65	10.83 ± 12.31	14.90 ± 12.31
Move 2	5 ± 0	61.83 ± 43.18	80.40 ± 114.75	92.29 ± 73.73
Move 3	10.67 ± 26.94	12.35 ± 18.26	7.35 ± 6.24	7.84 ± 7.40
Move 4	7.60 ± 8.26	9.42 ± 11.30	10.61 ± 11.90	7.99 ± 7.99
Move 5	43.81 ± 46.82	48.71 ± 44.46	26.63 ± 38.10	38.32 ± 58.75
Move 6	37.93 ± 28.35	52.40 ± 51.60	31.43 ± 48.74	19.70 ± 19.52
Move 7	18.90 ± 26.21	15.20 ± 17.93	14.17 ± 16.19	17.60 ± 20.09
Move 8	7.95 ± 7.48	17.15 ± 18.65	9.11 ± 9.50	13.31 ± 27.61
Move 9	12.05 ± 10.68	17.24 ± 18.37	30.78 ± 28.37	11.82 ± 6.63

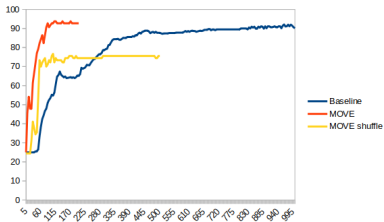
- Average length, with standard deviation, by move category, in tokens
- Column headers :
  - TA : Travel agencies
  - TB : Travel blogs
  - PTI : Public tourism information
  - POI : Points of interest

# Experimental protocol



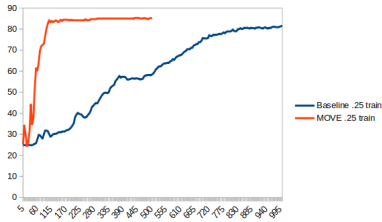
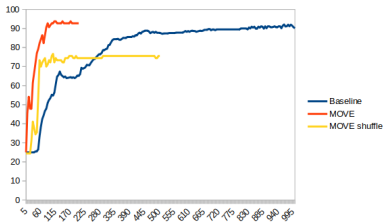
# Results

- Horizontal : epochs
- Vertical : Accuracy on the test test
- Hyperparameters tuned on the baseline



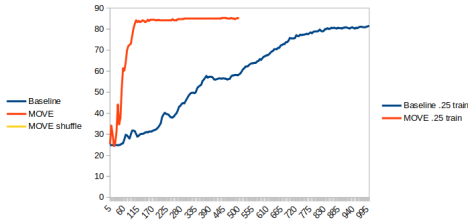
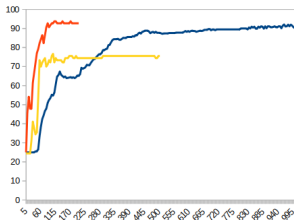
# Results

- Horizontal : epochs
- Vertical : Accuracy on the test test
- Hyperparameters tuned on the baseline



# Results

- Horizontal : epochs
- Vertical : Accuracy on the test test
- Hyperparameters tuned on the baseline



- High performance attained a lot more quickly : 100 epochs for MOVE and 500 epochs for the baseline in the first set of experiments
- Shuffling obtains the worst performance, indicating the influence of the move types

**Genre complexity** : - When analyzing genre, we have only a partial view of all that is really going on (Johns et al., 2006). The ways in which we analyze genre are only partial representations of the complex nature of genre and the social and communicative functions they have to fulfill.

- Analyzing digital discourse in tourism : we need to take a multidimensional approach, paying attention not only to the text, but also to other elements that are linked to the text (hyperlinks, visual graphic elements, but also other non-textual yet discursive and commercial elements).

- Web genre typology** : - Askehave and Nielsen's framework gap : it does not provide a clear explanation of how the concept of move might be operationalized in a text containing hyperlinks, which could interrupt a traditionally defined rhetorical unit of move (Mehlenbacher, 2017).
- Analysts may need to consider that digital genres offer readers different reading paths (Baldry et al., 2006) and that a reader may choose to read according to a dominant mode used in a text (Kress, 2003).



## **Annotation process :**

- 2 main problems - the establishment of the annotation guide based on Askehave and Nielsen (2005)
- the manual annotation : the text segmentation and a series of "borderline cases" where the segmented text can belong to two moves, which point to a certain degree of subjectivity.

- Information on moves seems highly informative
- The corpus is available (link in the paper)
- Future work :
  - Automatic move type identification as a task of interest
  - Refine the typology
  - Add multimodality (e.g. images)