

Présentation du jeu de données : Transcriptions de visites guidées au Palais des Beaux-Arts de Lille

Ce jeu de données est constitué de transcriptions des interactions lors de 4 visites guidées qui ont eu lieu au Palais des Beaux-Arts de Lille en 2022. Ces transcriptions ont été réalisées à partir de l'enregistrement audiovisuel des visites via la dispositif ci-dessous, qui mobilise plusieurs caméras et micro-cravates.



Pour plus d'informations sur le dispositif et la méthodologie, voir (Thiburce, 2024) : <https://books.openedition.org/irhis/8866>

Ces observations filmées ont été menées dans le cadre du projet de recherche *Augmented Artwork Analysis* financé par l'ANR et de FNR qui fait travailler ensemble la sémiotique, l'histoire de l'art et l'informatique, dans la conception d'un prototype d'application utilisée sur tablette numérique dans le musée : <https://anr.fr/Projet-ANR-20-CE38-0017>

Structure et confidentialité des données

Format et organisation : Le corpus est constitué de 4 fichiers PDF distincts, avec deux visites où les guides mobilisent une tablette numérique pour leur médiation et deux visites sans tablette numérique.

3 fichiers ont été nettoyés après transcription automatique via Whisper, outil mis à disposition sur la plateforme Huma-Num (ShareDocs) :

PBA_Lille_220317_AM_Sans_Tablette ; PBA_Lille_220519_AM_Avec_Tablette ;
PBA_Lille_220519_PM1_Avec_Tablette ;

Ces trois transcriptions adoptent des conventions qui consistent à rester au plus près des paroles des participants, en conservant les redoublements, les hésitations, etc. Pour ces transcriptions, il n'y a pas de ponctuation.

Un fichier n'a pas été nettoyé après transcription automatique via Whisper, outil mis à disposition sur la plateforme Huma-Num (ShareDocs) : PBA_Lille_220317_PM_Sans_Tablette

Déroulement des visites (salles et œuvres commentées)

Pour chaque visite, nous renseignons ci-dessous les salles parcourues et les tableaux commentés (avec temps de début, temps de fin et durée de médiation).

PBA_Lille_220317_PM_Sans_Tablette

Salles

Nr	Annotation	Temps de d...	temps de fin	Durée
1	École du Nord - XVIe + 1	00:03:24.000	00:42:30.000	00:39:06.000
2	Anvers - Rubens - Lille + 2	00:43:20.500	01:01:44.500	00:18:24.000
3	Chardin et la manière française +6	01:03:01.000	01:19:10.000	00:16:09.000
4	Hollande, Siècle d'or +15	01:20:04.040	01:24:50.040	00:04:46.000

Œuvres

Nr	Annotation			
1	Dirk Bouts	L'ascension des élus et La Chute des Damnés	00:03:24.000	00:12:13.000
2	Peter Brueghel Le Jeune	Le Dénombrement de Bethléem	00:12:14.000	00:29:18.000
3	Jan Van Hemessen	Vanité	00:29:45.000	00:42:30.000
4	Peter Paul Rubens	La Descente de Croix	00:43:20.500	01:01:44.500
5	Jean-Baptiste Siméon Chardin	Le gobelet d'argent	01:03:01.000	01:19:10.000
6	Abraham Von Beyeren	Nature morte au citron pelé et au verre	01:20:04.040	01:24:50.040

PBA_Lille_220317_AM_Sans_Tablette

Salles

Nr	Annotation	Temps de d...	temps de fin	Durée
1	École du Nord - XVIe + 1	00:09:30.000	00:25:17.000	00:15:47.000
2	Chardin et la manière française +6	00:26:23.000	00:31:52.000	00:05:29.000
3	Cabinet espagnol, Goya, Greco +13	00:34:49.000	00:44:15.000	00:09:26.000
4	France, Italie, Espagne 12e-15e s. -2	00:51:30.000	01:06:09.000	00:14:39.000

Œuvres

1	Jacob Jordaens	L'enlèvement d'Europe	00:09:30.000	00:25:17.000
2	Jean-Marc Nattier	La Duchesse de Lambesc et son frère le comte de Brienne	00:26:23.000	00:31:52.000
3	Francisco de Goya y Lucientes	Les Vieilles	00:34:49.000	00:44:15.000
4	Denis Foyatier	Spartacus brisant les liens	00:51:30.000	01:00:15.000
5	François Milhomme	Hermaphrodite	01:00:19.000	01:06:09.000

PBA_Lille_220519_AM_Avec_Tablette

Salles

Nr	Annotation	Temps de d...	temps de fin	Durée
1	École du Nord - XVIe + 1	00:04:15.000	00:34:19.000	00:30:04.000
2	Atrium, Espace Gigapixels	00:35:44.000	01:01:58.000	00:26:14.000

Œuvres

1	P. Brueghel Le Jeune	Le Dénombrement de Bethléem	00:04:15.000	00:34:19.000
2	P. Brueghel Le Jeune	Le Dénombrement de Bethléem	00:35:44.200	01:01:58.000

PBA_Lille_220519_PM1_Avec_Tablette

Salles

Nr	Annotation	Temps de d...	temps de fin	Durée
1	Courbet et le Réalisme + 9	00:00:00.000	00:55:11.384	00:55:11.384
2	Le Paysage +10	00:55:11.384	01:00:37.000	00:05:25.616
3	Académisme / Symbolisme / Esprit de Salon +11	01:00:37.500	01:10:39.240	00:10:01.740

Œuvres

1	Gustave Courbet	L'Après-dînée à Ornans	00:00:00.000	00:22:06.896
2	Jean François Millet	La Becquée	00:22:25.396	00:38:01.000
3	Jules-Adolphe Breton	Plantation d'un calvaire	00:38:01.178	00:40:39.644
4	Albert Anker	Dans les bois	00:41:01.600	00:43:07.411
5	Louis-Émile Salomé	La Maison de Thérèse	00:43:07.852	00:43:23.205
6	Amand Gautier	La promenade des soeurs	00:44:36.823	00:46:04.176
7	Carolus-Duran (Charles Auguste Émile Durant)	L'assassiné	00:47:24.137	00:49:47.230
8	Carolus-Duran (Charles Auguste Émile Durant)	L'homme endormi	00:49:47.230	00:52:24.076
9	Amand Gautier	Le Docteur Paul Gachet	00:52:25.692	00:53:24.461
10	Carolus-Duran (Charles Auguste Émile Durant)	La Dame en noir	00:53:24.807	00:54:40.846
11	Carolus-Duran (Charles Auguste Émile Durant)	Gustave Tempelaere (1840-1904), marchand de tableaux à Paris, ...	00:54:41.345	00:55:06.999
12	Constant Troyon	En forêt de Fontainebleau	00:55:23.000	00:58:12.153
13	Théodore Rousseau	La Seine à Villeneuve-Saint-Georges	00:58:26.809	00:59:26.230
14	Henri-Joseph-Constant Dutilleul	Hétraie dans la forêt de Fontainebleau	00:59:26.688	01:00:34.923
15	Carolus-Duran (Charles Auguste Émile Durant)	Le baiser	01:00:56.499	01:03:33.189
16	Carolus-Duran (Charles Auguste Émile Durant)	La Dame au chien / Madame Georges Petit	01:03:33.690	01:04:10.081
17	Carolus-Duran (Charles Auguste Émile Durant)	Madame Georges Petit	01:04:10.337	01:05:43.153
18	Carolus-Duran (Charles Auguste Émile Durant)	La Dame au chien	01:05:43.461	01:09:19.520
19	Edouard Manet	Berthe Morisot à l'éventail	01:09:19.781	01:10:31.911

Anonymisation : Toutes les données ont été anonymisées afin de garantir la confidentialité et le respect de la vie privée des étudiants. Aucun élément nominatif ou permettant l'identification des auteurs n'est présent dans les fichiers mis à disposition.

Des pseudonymes ont été utilisés pour identifier les participants : GUI = guide ; EN1 = Enseignant ; VI1 = visiteur 1 ; etc.

Pistes d'exploitation pour le hackathon

Ces transcriptions permettraient de travailler sur différentes dimensions du langage en interaction, notamment sur les variations entre grammaire de l'écrit et grammaire de l'oral.

Plusieurs pistes pourraient être travaillées lors de ce hackathon :

- les formes d'adresse entre les guides et les participants, les dynamiques de points de vue individuels et collectifs : repérer et caractériser les différents emplois des pronoms personnels, identifier leurs rôles au fil de la visite (par exemple, un « on » qui exprime le point de vue de la doxa ; un « on » qui renvoie à une instance anonyme ; un « on » qui réfère aux personnes en présence ; etc.) ;
- les dynamiques de référence aux œuvres croisées au fil des visites, les anticipations et les reprises d'éléments qui ont été vus, montrés et commentés dans le parcours ;
- l'expression de la perception des visiteurs et les marqueurs de subjectivité, les modes d'intégration des publics dans le discours des guides ;

Intérêt scientifique

Les transcriptions des échanges lors des visites guidées donnent accès à des discours situés sur les œuvres, face aux œuvres. Elles peuvent constituer un objet particulièrement pertinent d'un point de vue interdisciplinaire, à la croisée des sciences du langage, de l'histoire de l'art, de la sociologie et de l'informatique. En tant que traces d'interactions entre guides et publics, elles permettent d'interroger les relations qui se font jour entre perception et connaissances des œuvres, entre l'espace-temps de l'expérience des œuvres et l'histoire dans laquelle elles s'insèrent, le patrimoine auquel elles appartiennent.

Les participants au hackathon pourront ainsi explorer la capacité des grands modèles de langage à :

- repérer et caractériser différentes unités linguistiques ;
- faire émerger des régularités voire des patterns discursifs ;
- représenter les dynamiques d'interaction entre guides et publics ;
- pointer les potentialités et les limites des outils de l'IA face au langage oral et à la rhétorique investie dans le discours de médiation.