

# Projet AGODA

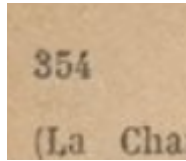
## Guide annotation de l'OCR

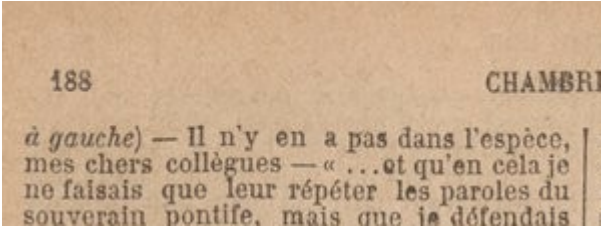
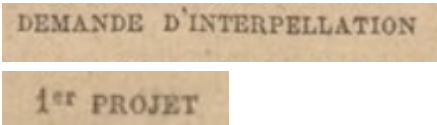
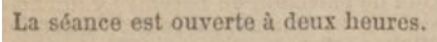
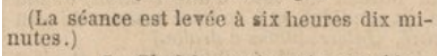
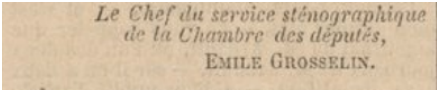
Fanny Lebreton et Marie Puren

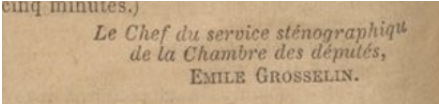
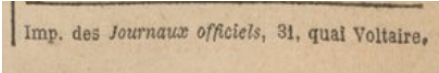
<b>Structure des débats</b>	<b>1</b>
Grandes divisions des débats	1
Paragraphes	3
Prises de parole	6
Cas particuliers	7
<b>Éléments spécifiques à la première page</b>	<b>10</b>
<b>Annexes</b>	<b>13</b>

## Structure des débats

### Grandes divisions des débats

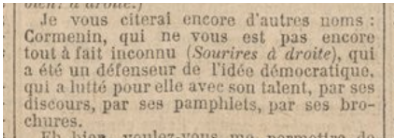
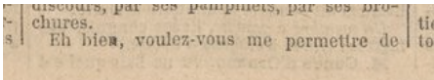
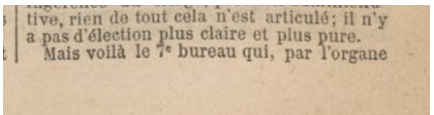
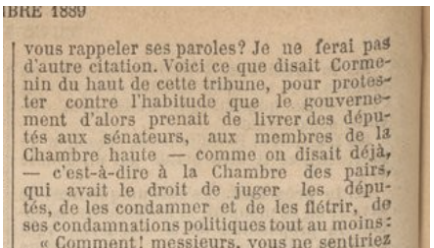
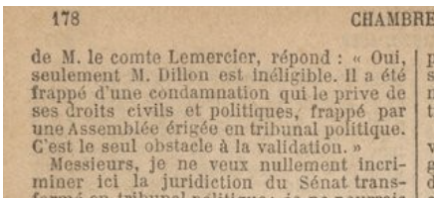
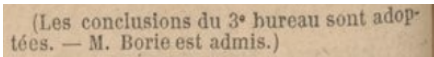
Tags	Utilisation	Balises TEI	Exemples
page-number	Annotation du numéro de page	<pb n="354"/>	

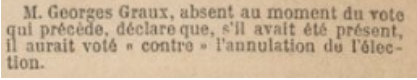
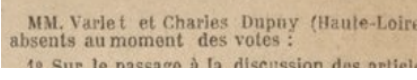
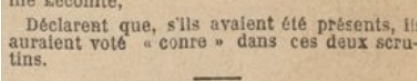
page-number-ref	Annotation d'un numéro de page dans incident	<pre>&lt;incident&gt;   &lt;desc&gt;     &lt;ref target="#p123"/&gt;   &lt;/desc&gt; &lt;/incident&gt;</pre>	
part head	Titre d'une division d'un débat	<pre>&lt;/div&gt; &lt;div type="part"&gt; &lt;head&gt;text&lt;/head&gt;</pre>	
opening seg	Texte signalant l'ouverture de la séance	<pre>&lt;note type="opening"&gt; &lt;seg&gt;La séance est ouverte à deux heures.&lt;/seg&gt; &lt;/note&gt;</pre>	
closing seg	Texte signalant la fin de la séance	<pre>&lt;note type="closing"&gt; &lt;seg&gt;(La séance est levée...)&lt;/seg&gt; &lt;/note&gt;</pre>	
signed seg back	Signature du chef du service sténographique et début des parties annexes	<pre>&lt;signed&gt; &lt;seg&gt;Le Chef du service sténographique de la Chambre des députés.&lt;/seg&gt; &lt;/signed&gt; &lt;/div&gt; &lt;/div&gt; &lt;/body&gt; &lt;back&gt;</pre>	

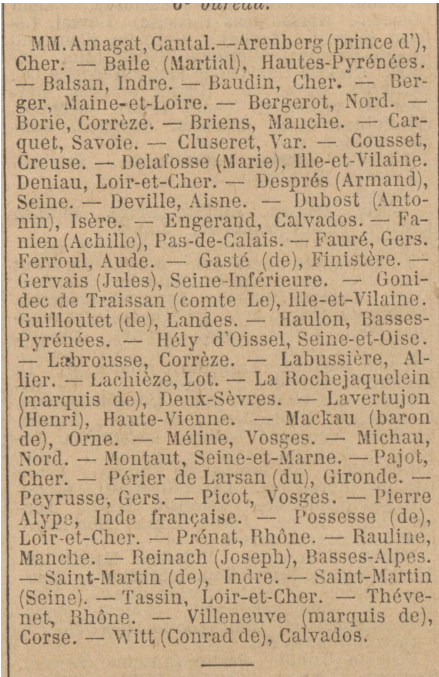
signed seg div-end	Fin d'une séance suivant une séance précédente	<pre>&lt;signed&gt; &lt;seg&gt;Le Chef du service sténographique de la Chambre des députés.&lt;/seg&gt; &lt;/signed&gt; &lt;/div&gt;</pre>	
text	Fin du débat sans annexe	<pre>text &lt;/div&gt; &lt;/body&gt; &lt;/text&gt;</pre>	
text-back	Fin du débat avec annexe	<pre>text &lt;/div&gt; &lt;/div&gt; &lt;/back&gt; &lt;/text&gt;</pre>	
leave	Excuses et demandes de congés	div de type part avec un attribut corresp qui correspond à l'étiquette	

## Paragraphes

Les interventions des députés sont souvent longues et divisées en paragraphes dans le *Journal Officiel*.

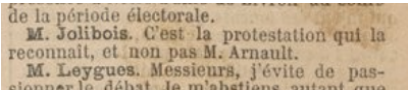
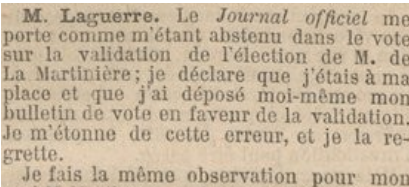
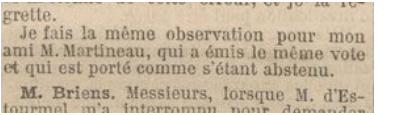
Tags	Utilisation	Balises TEI	Exemples
seg	Un paragraphe complet, non coupé par un changement de page ou de colonne	<seg>text</seg>	
seg-beginning	Début d'un paragraphe coupé par un changement de colonne	<seg>text	
	Début d'un paragraphe coupé par un changement de page		
seg-end	Fin d'un paragraphe qui a débuté à la fin de la colonne précédente et qui se poursuit au début de la colonne suivante	text</seg>	
	Paragraphe qui a débuté à la fin de la page précédente et se poursuit au début de la page suivante.		
comment seg	Paragraphe qui ne correspond pas à une division d'une prise de parole mais apportant une information sur le fonctionnement	<note type="comment"> <seg>Texte</seg> </note>	

Tags	Utilisation	Balises TEI	Exemples
	de la Chambre (texte entre parenthèses)		
	Paragraphe qui ne correspond pas à une division d'une prise de parole mais correspondant à un propos rapporté		
comment-beginning seg	Premier paragraphe d'une suite de paragraphes correspondant à un commentaire.	<note type="comment"> <seg>Texte</seg>	
comment-end seg	Dernier paragraphe d'une suite de paragraphes dans une note.	<seg>Texte</seg> </note>	

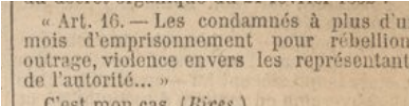
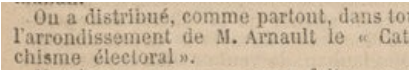
Tags	Utilisation	Balises TEI	Exemples
note seg	Paragraphe unique ne correspondant ni à une division dans une prise de parole, ni à un commentaire	<note> <seg>texte</seg> </note>	

## Prises de parole

Une prise de parole encodée avec la balise <u> correspond au moment où un orateur prend la parole de manière continue, sans interruption.

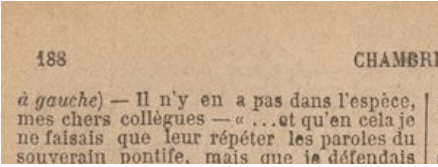
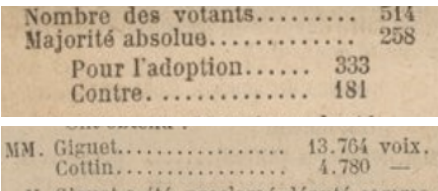
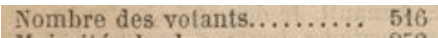
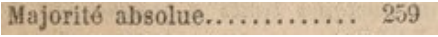
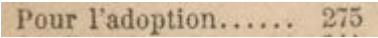
Tags	Utilisation	Balises TEI	Exemples
u seg	Prise de parole correspondant à un seul paragraphe.	<u> <seg>text</seg> </u>	
u-beginning seg	Premier paragraphe d'une prise de parole qui se poursuit sur un paragraphe ou plus	<u> <seg>text</seg>	
u-end seg	Dernier paragraphe d'une prise de parole qui a débuté avec un ou plusieurs paragraphes	<seg>text</seg> </u>	

## Cas particuliers

Tags	Utilisation	Balises TEI	Exemples
quote seg	Paragraphe entre guillemets correspondant à une citation	<seg> <quote>text</quote> </seg>	
	Citation apparaissant dans un paragraphe	<seg>texte <quote>text</quote> </seg>	

quote-beginning seg	Début d'une citation marqué par un guillemet ouvrant	<quote>texte	On ne se borne pas à dire que l'aïeul de M. Cambe s'était installé sans droit dans le domaine de Cornusson, qu'il s'en était attribué la propriété par des moyens frauduleux, ce qui était déjà très grave. Non! on ajoutait ceci pour combler la mesure : « Au sujet de votre aïeul, je me permettrai de vous poser quelques ques-
quote-end seg	Fin d'une citation marquée par un guillemet fermant	texte</quote>	« N'est-il pas vrai que votre grand-père était membre du comité de salut public à Villefranche-d'Aveyron, avec Joseph Mollinier, d'Arnac, qui trouvait bon de garder les calices, les encensoirs en argent, qu'il s'empressait de vendre à son profit; avec Verdié et Mouly, qui vendaient la liberté des citoyens à un prix déterminé, infligeant de nouveau la reclusion à ces mêmes citoyens, dont ils extorquaient une somme pour les rendre à la liberté; qui volaient, opportunistes de ce temps néfaste, pillaient, commettaient les atrocités les plus révoltantes... »
incident	Texte correspondant à la description d'un événement perturbant le débat (bruits, cris, protestations) - entre parenthèses et en italique <i>*Il peut y avoir plusieurs incidents dans un seul paragraphe.</i>	<incident> <desc>(Texte)</desc> </incident>	M. Cuneo d'Ornano. ... le pouvoir absolu, je le répète. (Non! non! au centre.) Vous n'êtes arrêtés ni gênés par aucune considération d'aucun ordre, sauf celles qui vous paraîtront dictées par votre raison. Voyons! Ces noms-là ne vous disent plus rien aujourd'hui? (Exclamations sur les mêmes bancs.)
incident-beginning	Incident qui se poursuit à la page suivante	<incident> <desc>(Texte	« Le dimanche suivant, 15 septembre, après avoir annoncé les fêtes de la semaine, je dis à mes paroissiens que je regrettais que certains eussent mal interprété mes paroles; que j'avais voulu leur dire uniquement de s'éloigner des ennemis de l'Eglise et en particulier des francs-maçons...» (Bruit



incident-end	Incident qui a débuté sur la page précédente	Texte)</desc> </incident>	
table	Traitement des tableaux	<table> <row> <cell>Tout le texte du tableau</cell> </row> </table>	
other-sitting	Début d'une autre séance au sein de la même séance (cf. séance du 14 janvier 1890)	</div> <div type="other-sitting"> <head>2ème séance</head>	
nb-voters	Résultats d'un vote dans le texte des débats : ligne correspondant au nombre de votants	<desc> <measure type="nbvoters">Texte</measure>	
majority	Résultats d'un vote dans le texte des débats : ligne correspondant à la majorité absolue	<measure type="majority">Texte</measure>	
ayes	Résultats d'un vote dans le texte des débats : ligne correspondant aux nombres de votes "pour"	<measure type="ayes">Texte</measure>	

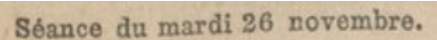
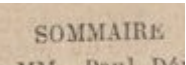
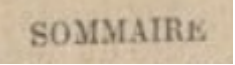

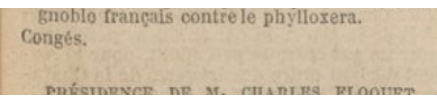
noes	Résultats d'un vote dans le texte des débats : ligne correspondant aux nombres de votes "contre"	<measure type="noes">Texte</measure> </desc>	Contre..... 241
------	--	---	-----------------

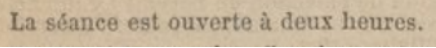
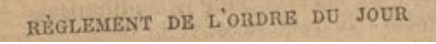
## Éléments spécifiques à la première page

La première page d'une séance de débats parlementaires présente des éléments uniques, dans une mise en page spécifique. Ces éléments requièrent donc d'utiliser des tags spécifiques

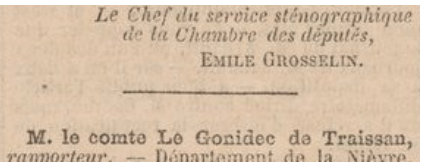
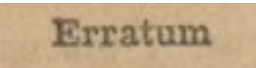
Tags	Utilisation	Balises TEI	Exemples
body page-number date-pub	Annotation du numéro de la première page.  <i>*Dans la boîte, on aura le texte suivant : "Journal officiel du 27 novembre 1889 CHAMBRE - SÉANCE DU 26 NOVEMBRE Session extraordinaire de 1889 175." L'élément qui nous intéresse ici est le numéro de page "175".</i>	<text> <body> <pb n="175"/>	Journal officiel du 27 Novembre 1889 CHAMBRE - SÉANCE DU 26 NOVEMBRE Session extraordinaire de 1889 175
	Extraction et encodage de la date de publication du numéro du journal renseignée dans le <date> de <biblfull>	<sourceDesc> <biblfull> <titleStmt> ... </titleStmt>	

Tags	Utilisation	Balises TEI	Exemples
	<p><i>*Dans la boîte, on aura le texte suivant : "Journal officiel du 27 novembre 1889 CHAMBRE - SÉANCE DU 26 NOVEMBRE Session extraordinaire de 1889 175."</i></p> <p><i>L'élément qui nous intéresse ici est la date de publication "27 novembre 1889".</i></p>	<pre>&lt;publicationStmt&gt; &lt;publisher&gt; ... &lt;/publisher&gt; &lt;pubPlace&gt; ... &lt;/pubPlace&gt; &lt;date&gt; 27 novembre 1889&lt;/date&gt; &lt;publicationStmt&gt; &lt;date&gt;</pre>	
body1 page-number	Indique qu'il s'agit de la page n°1 de la nouvelle session		
useless	Texte non utilisé dans le <text>		CHAMBRE DES DÉPUTÉS
meeting-session meeting-legislature	Indication du type de la session. La valeur de l'attribut @n est égale à la première lettre en majuscule du mot suivant le terme "Session" et du numéro de session.	<pre>&lt;meeting n="E1" ana="#parla.lower #parla.session"&gt;Session extraordinaire de 1889&lt;/meeting&gt;</pre>	5 <sup>e</sup> législature. — Session extraordinaire de 1889.
	Indication de la législature. La valeur de l'attribut @n est égale au chiffre débutant la ligne suivi de la lettre L	<pre>&lt;meeting n="5L" ana="#parla.lower #parla.legislature"&gt;5e législature&lt;/meeting&gt;</pre>	
meeting-sitting	Indication du numéro de séance. La valeur de l'attribut @n est	<pre>&lt;meeting n="10" ana="#parla.lower</pre>	COMPTE RENDU IN EXTENSO. — 10 <sup>e</sup> SÉANCE

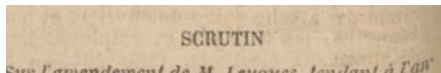
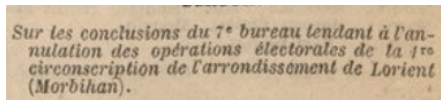
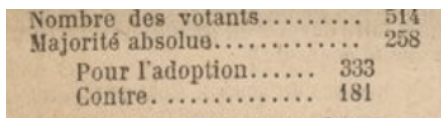
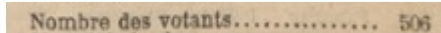
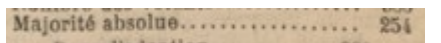
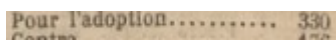

Tags	Utilisation	Balises TEI	Exemples
	égale à au nombre précédant le "e" en indice.	#parla.sitting">10e séance</meeting>	
date-sitting	Date de la séance	<settingDesc> <setting> .... <date>mardi 26 novembre<date> </setting>	
sitting	Début de la séance  <i>En général, il n'y a qu'une seule séance par numéro et rien n'indique, dans la présentation, où celle-ci débute. Nous considérons donc que celle-ci débute avant le sommaire et qu'il s'agit de la 1ère séance. En effet, celle-ci peut-être suivie d'une seconde séance matérialisée par un titre (cf. other-sitting)</i>	<div type="sitting">	
contents	Titre du sommaire	<div type="contents"> <head> SOMMAIRE </head> <list>	
item	Un élément du sommaire	<item>text</item>	
item-list	Dernier élément du sommaire	<item>text<item> </list>	

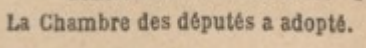
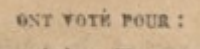
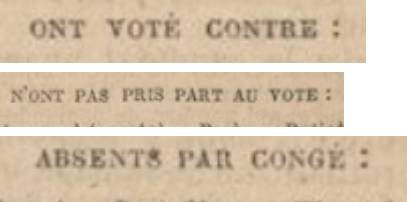
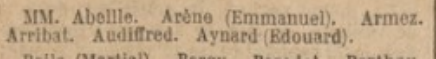

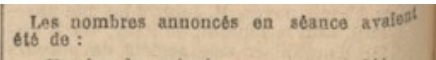
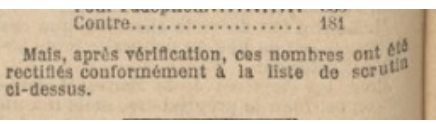
Tags	Utilisation	Balises TEI	Exemples
opening seg	Texte signalant l'ouverture de la séance	<note type="opening"> <seg>La séance est ouverte à deux heures.</seg> </note>	
agenda	Titre de la partie consacrée au règlement de l'ordre du jour	</div> <div type="agenda"> <head>Règlement de l'ordre du jour</head>	

## Annexes

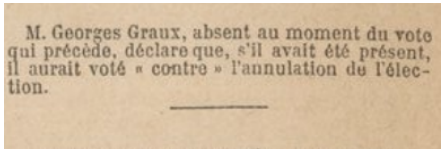
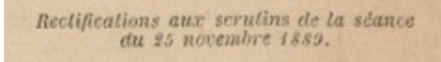
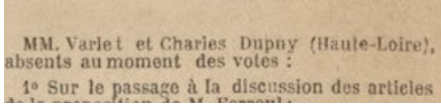
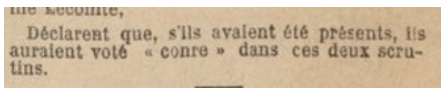
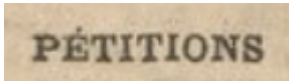
Tags	Utilisation	Balises TEI	Exemples
part1(-XXXX)	Indique qu'il s'agit de la première division de l'annexe et qu'il est donc inutile d'ajouter une </div> avant la <div> indiquant le début d'une partie	<div type="part">	
	S'il s'agit d'une division typée, on ajoute le type après le tiret. Par exemple : part1-erratum, part1-lists, etc.	<div type="erratum">	

Tags	Utilisation	Balises TEI	Exemples
part1 u-beginning seg	Première division de l'annexe sans titre (après la signature)	<div type="part"> <u> <seg>texte</seg>	
erratum	Début de l'erratum ou des errata	</div> <div type="erratum"> <head> <label>Erratum</label>	
note-head	Précision portant sur le titre d'une partie	<note>texte</note> </head>	
lists	Début des listes	</div> <div type="lists"> <head> <label>Listes</label>	
offices	Début de la nomination des bureaux	</div> <div type="offices"> <head>Texte</head>	
appendices	Début "officiel" des annexes	</div> <div type="appendices"> <head>text</head>	
voting1	Début du 1er vote	<div> <div type="voting"> <head> <label>SCRUTIN</label>	

Tags	Utilisation	Balises TEI	Exemples
voting	Début des votes suivants	</div> <div type="voting"> <head> <label>SCRUTIN</label>	
note-head	Informations sur l'objet du vote	<note>Texte</note> </head>	
table	Traitement des tableaux	<table> <row> <cell>Tout le texte du tableau</cell> </row> </table>	
nb-voters	Nombre de votants	<desc> <measure type="nbvoters">Texte</measure >	
majority	Majorité absolue	<measure type="majority">Texte</measure >	
ayes	Nombres de votes "pour"	<measure type="ayes">Texte</measure>	
noes	Nombres de votes "contre"	<measure type="noes">Texte</measure> </desc>	

Tags	Utilisation	Balises TEI	Exemples
result	Résultat du vote	<note type="result"> <seg>Texte</seg> </note>	
voterslist-beginning	Début de la liste des votants	<note type="voterslist"> <desc>texte</desc>	
desc	Titre des différents résultats du vote	<desc>Texte</desc>	
seg	Paragraphe correspondant à un groupe de noms de votants	<seg>Texte</seg>	
seg note-end	Dernier paragraphe des noms des votants	<seg>Texte</seg> </note>	
seg note-beginning	Premier paragraphe récapitulant les résultats du vote	<note type="numbersannounced"> <seg>Texte</seg>	
seg note-end	Dernier paragraphe d'une partie consacrée à un vote	<seg>Texte</seg> </note>	



Tags	Utilisation	Balises TEI	Exemples
seg note-end div-end	Dernier paragraphe qui clôt la série des votes	<note> <seg>Texte</seg> </note> </div>	
rectification	Titre de la partie consacrée aux rectifications	</div> </div> <div type="rectification"> <head>texte</head>	
seg comment-beginning	Premier paragraphe d'une note lorsque celle-ci est divisée en plusieurs paragraphes	<note type = "comment"> <seg>Texte<seg>	
seg note-end	Dernier paragraphe d'une note lorsque celle-ci est divisée en plusieurs paragraphes	<seg>Texte</seg> </note>	
petition	Titre de la partie consacrée aux pétitions	</div> <div type="petition"> <head> <label>text</label>	
note-head	Informations concernant l'objet des pétitions	<note>Texte</note> </head>	