

Les données utilisées par Ariane

Ariane écoute une séance publique et reconstitue au fil de l'eau qui prend la parole, sur quel amendement, avec quel résultat de vote. Encore faut-il choisir avec soin les données que le système a le droit de regarder. Ce document explique lesquelles, et pourquoi les autres sont écartées.

Deux familles de données

L'Assemblée nationale publie beaucoup de données autour d'une séance. Elles se répartissent en deux familles. Les unes sont des listes stables — députés en exercice, amendements déposés, textes examinés — connues **avant** la séance. Les autres — comptes rendus, votes détaillés, amendements consolidés — ne paraissent qu'**après** le débat, une fois le travail de rédaction achevé. Précieuses, elles n'existent pourtant pas encore quand la séance se déroule : elles ne peuvent pas alimenter un système qui travaille en direct.

Le déroulement du débat lui-même, minute par minute, n'est couvert par aucune publication. Les seules données qui circulent pendant la séance viennent des outils de travail de l'Assemblée.

Le critère de tri : une entrée ne dit jamais qui parle

Une source n'est une **entrée** utile que si elle remplit trois conditions : elle est produite par une machine, elle est disponible pendant la séance, et elle ne dit pas qui est en train de parler.

La troisième condition est la plus importante. Dès qu'une source nomme l'orateur, c'est qu'un être humain l'a saisie en régie — le poste depuis lequel des agents de l'Assemblée suivent la séance et notent au clavier, en direct, qui prend la parole. Or c'est exactement le travail qu'Ariane cherche à automatiser. Une telle source n'est donc pas une entrée : c'est une **sortie**, le résultat attendu et non la matière première. S'en servir pour construire le fil reviendrait à recopier la réponse.

Rôle	De quoi il s'agit
Entrées	Les sources produites par machine, pendant la séance, qui donnent le contexte sans nommer l'orateur.
Sortie	Le fil produit par Ariane : qui parle, à quel moment, relié aux fiches officielles.
Version de référence	Le sommaire saisi par la régie, utilisé après coup pour mesurer si Ariane a vu juste.

Les sources utilisées pendant la séance

Trois entrées seulement, décrites ici par leur nature :

- **Le programme de séance** (que l'Assemblée appelle le « dérouleur ») : la trame des articles et des amendements appelés, avec les identifiants officiels des textes, des amendements et de leurs auteurs, et l'annonce des scrutins publics — ces votes annoncés à l'avance où la position de chaque député est enregistrée.
- **Le suivi des amendements** (Eliasse, l'outil interne que l'Assemblée utilise pendant le débat pour afficher où en est l'examen) : l'amendement en cours ou sur le point d'être discuté, et son sort en séance — adopté, rejeté, tombé (devenu sans objet) ou retiré.

- **Le son et la vidéo** de la séance : les seules données qui existent en même temps que le débat, mais aussi les seules où rien n'est identifié à l'avance — ni le nom de l'orateur, ni le numéro de l'amendement. Un flux de paroles, et rien d'autre.

Ariane enregistre ces trois entrées en notant l'heure exacte de chaque élément, ce qui permet de rejouer ensuite la séance à l'identique. Elle enregistre aussi, à part, le sommaire saisi par la régie — non pour alimenter le fil, mais pour pouvoir le comparer après coup.

Ces trois entrées disent le *quoi* : quel texte, quel article, quel amendement, quel auteur d'amendement. Aucune ne dit le *qui* : la personne au micro n'est pas forcément l'auteur de l'amendement — ce peut être le rapporteur, un ministre, le président de séance, ou un député qui fait un rappel au règlement. Produire cette attribution en direct, à partir de la reconnaissance vocale, c'est le travail d'Ariane. Même logique pour les votes : l'annonce d'un scrutin et son issue sont connues pendant la séance, mais le détail chiffré n'est publié qu'après.

Du nom entendu à la fiche officielle

Un nom reconnu à l'oreille ne suffit pas : encore faut-il le rattacher à la bonne personne. C'est le rôle des référentiels, les listes officielles et publiques de l'Assemblée — députés en exercice, amendements déposés, documents examinés. Elles changent peu et attribuent à chaque député et à chaque amendement un identifiant stable, qui ne bouge pas d'une séance à l'autre. Ariane s'en sert uniquement pour retrouver l'identité exacte de ce qui a été *entendu*, jamais pour deviner à sa place. Comme elles changent peu, ces listes peuvent être copiées à l'avance : Ariane n'a pas besoin d'être connectée pour les consulter.

Reste le sommaire saisi par la régie, qui recense les orateurs effectifs, les suspensions et les rappels au règlement. Il n'entre jamais dans la fabrication du fil : il est mis de côté, puis comparé après coup au résultat d'Ariane, comme version de référence, pour mesurer l'écart. Il évalue le travail d'Ariane ; il ne l'alimente pas.