

# Académie française

~

Réalisation d'un nouveau site Internet pour la consultation du

# Dictionnaire de l'Académie française

dans sa 9<sup>e</sup> édition, ainsi que les 8 éditions précédentes

Partie II : Cahier des charges prévisionnel

décembre 2017 version 1.2

 $\sim$ 

Rédaction : Laurent Catach Société 4H Conseil laurent.catach@bbox.fr





# **Préambule**

L'Académie française, membre de l'Institut de France, publie depuis plus de trois siècles son *Dictionnaire* de l'Académie française (ci-après appelé le *Dictionnaire* ou le DAF), dont la 9<sup>e</sup> édition est en cours de rédaction et devrait se terminer à l'horizon fin 2019/2020<sup>1</sup>.

Afin de valoriser son *Dictionnaire* et, dans le cadre de sa mission de service public, de le rendre accessible à tous, l'Académie dispose déjà d'une version numérique de cet ouvrage, disponible en accès libre sur Internet<sup>2</sup>. Cette version est cependant trop ancienne, et à l'occasion de l'achèvement prochain de la dernière édition, l'Académie souhaite aujourd'hui pouvoir entamer dès à présent une **refonte** complète de la version numérique du DAF, afin de le rendre plus simple d'accès, plus moderne et plus riche en fonctionnalités.

Cette nouvelle version numérique prendra la forme d'un site Internet, dont la première version serait à paraître fin juin 2018.

Par ailleurs, l'Académie a acquis récemment les droits d'exploitation sur une base informatique comprenant l'ensemble des 8 éditions précédentes du *Dictionnaire*, depuis la 1<sup>re</sup> édition en 1694<sup>3</sup>. La seconde étape du présent projet sera ainsi d'**intégrer l'ensemble de ces éditions** au nouveau site Internet, avec des fonctionnalités supplémentaires pour la consultation et la navigation, offrant par là même à la communauté francophone et scientifique un corpus de grande qualité pour l'étude de la langue française.

# Contenu de ce document

Ce document constitue une pré-version simplifiée du cahier des charges, à destination des prestataires qui seront envisagés pour devenir partenaires du développement informatique du projet. Il décrit l'ensemble des fonctionnalités et des contenus éditoriaux prévus, mais sans rentrer dans les détails plus techniques. Des cahiers des charges plus précis seront réalisés ultérieurement, après les premières réunions de travail.

Ce document a donc pour but de permettre aux partenaires de réaliser un **devis** pour les prestations demandées. Dans ce cadre, certaines fonctionnalités pourront ou non être développées, selon la difficulté, les priorités et le savoir-faire du prestataire.

# **Prestations attendues**

Pour la réalisation de ce projet, il est attendu du prestataire la prise en charge des travaux suivants :

- une analyse détaillée et un retour sur le présent cahier des charges, selon l'expérience du prestataire, afin de définir conjointement la faisabilité et la pertinence des caractéristiques du projet, et de prioriser et planifier les développements sur plusieurs versions (2018-2019).
- le **développement initial** du site Internet, incluant la mise au point de processus d'indexation automatique des données, la mise en place d'un serveur de tests, l'intégration de l'interface, et bien sûr l'intégration des fonctionnalités souhaitées en version 1.
- l'hébergement du site Internet, en prévoyant notamment un dimensionnement adéquat du nouveau site (en termes de trafic), ainsi qu'une optimisation pour les moteurs de recherche (SEO).
- la mise en place d'un **backoffice**, même simple, permettant de gérer le site, d'en administrer certaines parties (comme la page d'accueil) et d'accéder aux statistiques.

<sup>&</sup>lt;sup>1</sup> Pour en savoir plus sur le *Dictionnaire* de l'Académie, voir le site <a href="http://academie-française.fr/le-dictionnaire/la-9e-edition">http://academie-française.fr/le-dictionnaire/la-9e-edition</a>. Les 3 premiers tomes ont été publiés, le 4<sup>e</sup> tome est en cours de rédaction.

<sup>&</sup>lt;sup>2</sup> Cette version numérique est hébergée par l'ATILF / CNRS. Le *Dictionnaire* est en fait accessible via deux interfaces différentes : <a href="http://atilf.atilf.fr/academie9.htm">http://atilf.atilf.fr/academie9.htm</a> et <a href="http://www.cnrtl.fr/definition/academie9/">http://www.cnrtl.fr/definition/academie9/</a>. Ces deux sites proposent également l'accès à la 8<sup>e</sup> édition. Voir également en annexe pour en savoir plus.

<sup>&</sup>lt;sup>3</sup> Les éditions du DAF : 1<sup>re</sup> (1694), 2<sup>e</sup> (1718), 3<sup>e</sup> (1740), 4<sup>e</sup> (1762), 5<sup>e</sup> (1798), 6<sup>e</sup> (1835), 7<sup>e</sup> (1878), 8<sup>e</sup> (1932-35). Le corpus de ces anciennes éditions a été racheté à la société Le Robert, qui avait acquis le fonds Redon au début des années 2000 (voir page 4 et en annexe).



• la **maintenance**, les **mises à jour** et les **évolutions** du site, via des correctifs ainsi que les évolutions du projet, dont les grandes lignes sont décrites ci-après.

Ce document doit permettre au prestataire d'établir des **devis** et des **plannings** pour les différentes facettes et versions du projet.

L'Académie souligne par ailleurs qu'elle souhaite s'engager sur le long terme dans ce projet, ce qui a pour conséquence que la capacité du prestataire à garantir la bonne collaboration, le suivi, la mise à disposition d'un **interlocuteur dédié** et la maintenance technique du nouveau site Internet sur la durée sont des éléments essentiels dans la réussite du projet.

L'Académie souhaite également que le prestataire puisse **être force de proposition** pour la conception initiale, les fonctionnalités et les évolutions du site, de manière à bénéficier d'un savoir-faire (technique, ergonomique, etc.), qui est en constante évolution sur Internet.

En termes de budget, l'Académie prévoit de dégager une enveloppe initiale pour le développement de la version 1 (qui représentera l'effort le plus important), puis un budget annuel plus réduit dédié aux coûts d'hébergement, de maintenance et d'évolutions du site.

# Planning des versions

Afin d'étaler les développements, de publier une première version dès 2018, et d'intégrer progressivement les données disponibles, le projet sera d'ores et déjà organisé autour de 3 étapes importantes :

- **version 1**: parution fin juin 2018; cette première version intègrera l'édition actuelle (mais incomplète) de la 9<sup>e</sup> édition du *Dictionnaire* (ci-après abrégée en « DAF9 »), ainsi que la 8<sup>e</sup> édition (DAF8) de manière à pouvoir proposer des articles pour tous les mots de A à Z<sup>4</sup>. Elle intègrera bien évidemment la nouvelle **interface** du site, un ensemble de fonctions de consultation considérées comme indispensables, ainsi que le nouveau module de **conjugaison**.
- **version 2:** parution fin 2018/début 2019; cette seconde version intègrera toutes les éditions du *Dictionnaire*, donnant ainsi accès à l'intégralité des 9 éditions. À cette occasion, une mise à jour de l'interface devra être effectuée, ainsi que l'ajout de nouvelles fonctionnalités, pour intégrer la consultation de ce nouveau corpus.
- **version 3 :** parution à date d'achèvement de la 9<sup>e</sup> édition du *Dictionnaire*, à l'horizon fin 2019/2020. À cette occasion, une mise à jour globale du site sera effectuée, selon les besoins du moment.
  - Le devisage de cette version 3 n'est pas à inclure pour la réponse au présent appel d'offres. Cette version est mentionnée dans le document uniquement pour anticiper des évolutions du projet.

Tous ces éléments sont décrits plus loin dans ce document.

Bien évidemment, des mises à jour correctives, ou bien des améliorations mineures, ne manqueront pas d'être effectuées entre chaque version, selon un planning à définir. Par ailleurs, le DAF9 paraissant à intervalles réguliers au *Journal Officiel*, sous forme de « cahiers » correspondant à des tranches alphabétiques, ces nouvelles données seront intégrées au fur et à mesure au site Internet, via des réindexations automatiques.

#### Sites de référence

Il existe de très nombreux sites dédiés à la consultation de dictionnaires, encyclopédies, conjugaisons, etc., certains en accès libres et d'autres à accès payants. De ce fait, il existe une forme de **standardisation** dans les usages pour ces type de corpus et de consultation (y compris via des smartphones), dont il est nécessaire de tenir compte afin d'être conforme à l'état de l'art et aux usages des internautes.

<sup>&</sup>lt;sup>4</sup> En juin 2018, on peut estimer que le DAF9 sera publié jusqu'à la lettre S. Les articles de T à Z seront donc proposés dans le DAF8.



On trouvera en annexe quelques sites de référence<sup>5</sup> (dont plusieurs anglo-saxons) dont il faudra s'inspirer pour la réalisation du présent projet, que ce soit sur le plan de l'interface, de l'ergonomie, des contenus, de la navigation, des fonctionnalités, de l'interopérabilité (API), etc. et également du point de vue de l'optimisation pour les moteurs de recherche (SEO)<sup>6</sup>.

Il existe par ailleurs déjà plusieurs sites proposant la consultation de certaines éditions du *Dictionnaire*, cités en annexe, et dont il conviendra également de faire une analyse (rapide) par le prestataire. Il a existé en particulier un CD-ROM, publié à la fin des années 1990 par les éditions Redon, intégrant les 8 premières éditions du Dictionnaire de l'Académie. Ce logiciel (PC) contient de nombreuses fonctionnalités tout à fait remarquables, ainsi que certaines idées d'interface originales, qui, bien que technologiquement dépassées, restent pertinentes pour le présent projet. De plus, les données exploitées par l'Académie sur ces 8 éditions proviennent précisément du fonds Redon. Il sera donc utile que le logiciel puisse également être réexaminé par le prestataire, en amont du projet.

De même, il existe déjà un site de consultation des 8 premières éditions du *Dictionnaire*, publié par les éditions Classiques Garnier numérique. Cependant, ce site n'est pas en accès libre, et son interface est vieillissante; il conviendra cependant d'être examiné pour le présent projet (voir en annexe).

# I. Contenus éditoriaux

Ce chapitre décrit les données qui seront mises en œuvre dans le site, leurs formats, et certains aspects concernant leur présentation. Les détails plus techniques ne sont pas donnés ici.

## I.1. Le Dictionnaire de l'Académie

En premier lieu, le site s'articulera bien sûr autour de la consultation du *Dictionnaire* de l'Académie française, qui sera au cœur du dispositif. Plus précisément, ce corpus sera constitué :

- de la 9<sup>e</sup> édition du *Dictionnaire*, en cours de rédaction, qui comprend la partie déjà publiée (de A à R) + la suite du dictionnaire (de S à Z), qui sera intégrée au fur et à mesure des publications<sup>7</sup>.
- du supplément (addenda) déjà publié de la 9<sup>e</sup> édition.
- de la 8<sup>e</sup> édition du dictionnaire, qui servira surtout à donner accès aux articles de S à Z, mais qui sera en réalité intégralement consultable.
- en version 2 : du corpus complet des 7 autres éditions du *Dictionnaire*.

Le format des données est indiqué plus loin dans ce document.

# I.2. Les nomenclatures

Véritable colonne vertébrale du site, la nomenclature générale du DAF sera le point d'entrée des **recherches** de mots dans le dictionnaire. Celle-ci :

• sera « développée », c'est-à-dire intègrera toutes les entrées mais aussi toutes les variantes, sous-entrées, etc. accessibles dans le DAF.

<sup>&</sup>lt;sup>5</sup> Le présent cahier des charges s'inspire en particulier des produits des Dictionnaires Le Robert, qui ont été créés par le rédacteur de ce document.

<sup>&</sup>lt;sup>6</sup> Les sites de dictionnaires gratuits génèrent habituellement de très forts trafics, et sont de ce fait très optimisés pour le référencement.

<sup>&</sup>lt;sup>7</sup> Un nouveau « cahier » paraît tous les 6 mois environ, avec une nouvelle tranche alphabétique, et ce jusqu'à l'achèvement fin 2019/2020.



- fusionnera toutes les nomenclatures des 9 éditions du DAF, en intégrant notamment les mots disparus de la 9<sup>e</sup> édition mais présents dans l'une des éditions précédentes; il s'agira donc d'une « supernomenclature ».
- permettra d'établir le lien entre les différentes éditions, sur un même mot, grâce à un système d'identifiants et des tables de correspondance.
- intègrera également de nombreuses métadonnées sur les mots, comme la catégorie grammaticale, les séries verbales, les dates de première entrée/sortie dans le DAF, etc.

À part les nomenclatures du *Dictionnaire*, ces bases contiendront également diverses autres nomenclatures ou index (par exemple la nomenclature des titres de notices, un index des proverbes et locutions, voir cidessous II.4), qui seront utilisées par le moteur de recherche du site, afin de collecter tous les résultats possibles dans les ressources disponibles.

# I.3. Les plans d'articles

Une base d'informations complémentaire sera fournie, correspondant aux **plans des articles longs** du *Dictionnaire*, sous la forme de chapitres (I., II....) avec des libellés. Ce système, baptisé « Explorateur », permettra d'afficher, pour ces articles, des « sommaires » interactifs, très utiles pour naviguer au sein d'un article long (voir plus loin).

# I.4. La base de conjugaison

Le nouveau site du DAF comprendra un nouveau module de **conjugaison** de tous les verbes du *Dictionnaire*, prenant appui sur une base de données de **formes fléchies** (verbes conjugués, ainsi que les pluriels et féminins, utilisés par ailleurs). Cette base contient plusieurs centaines de milliers de formes.

Ce module s'appuiera également sur un certain nombre de **métadonnées** sur les verbes (série verbale, choix de l'auxiliaire, etc.), appartenant à la nomenclature générale (ci-dessus). Les fonctionnalités de ce module sont décrites au II.5.

Cette base de formes fléchies joue un rôle central dans la consultation du site, puisqu'elle permettra également de mettre en œuvre la **navigation hypertexte** généralisée, les recherches en **texte intégral**, les **hyperappels** (appels externes du dictionnaire sur n'importe quel mot), les APIs, etc. Ces fonctionnalités sont décrites plus loin.

#### I.5. Les notices

Le nouveau site du DAF intègrera plusieurs ensembles de **notices**, de différents types mais d'un fonctionnement similaire :

- notices des rubriques « Dire, ne pas dire », publiées par l'Académie française<sup>8</sup>;
- notices des rubriques « Questions de langue » (idem)<sup>9</sup>;
- notices sur l'orthographe.

D'autres types de notices pourront être ajoutés à l'avenir.

Ces notices seront liées aux entrées du DAF par des liens hypertextes, activables soit par des hyperliens dans le texte des articles soit dans les « blocs d'informations » proposés en accompagnement de ceux-ci (voir II.1). Inversement, les notices renverront par des hyperliens aux articles du DAF.

Les notices seront mises en pages et consultables dans la même interface générale (HTML/CSS) que celle du *Dictionnaire*.

<sup>&</sup>lt;sup>8</sup> Voir <a href="http://academie-francaise.fr/dire-ne-pas-dire">http://academie-francaise.fr/dire-ne-pas-dire</a>.

<sup>&</sup>lt;sup>9</sup> Voir <a href="http://academie-francaise.fr/questions-de-langue">http://academie-francaise.fr/questions-de-langue</a>.



#### I.6. Les liens externes

Un certain nombre de **liens externes** seront proposés en accompagnement de la consultation du DAF, renvoyant à des ressources disponibles ailleurs sur Internet. Cela peut concerner par exemple la base France Terme<sup>10</sup>, le TLFi<sup>11</sup>, des ressources sur les régionalismes, etc.

En particulier, des liens seront proposés, en version 2, vers des **versions numérisées** (fac-similés, en mode PDF ou image) des anciennes éditions du *Dictionnaire*, comme par exemple sur Gallica<sup>12</sup>.

Ces liens pourront être soit générés automatiquement (à partir du mot-entrée, via une syntaxe ou API des sites concernés), soit issus d'une base de données éditorialisée. Comme précédemment, ils seront activables soit par des hyperliens dans le texte des articles soit dans les « blocs d'informations » proposés en accompagnement (voir II.1).

# I.7. Formats des données

Toutes les données textuelles seront fournies en **XML**, accompagnées d'une DTD/schéma. Ces données ont fait l'objet d'un important chantier de structuration, de balisage et de vérifications. Elles intègrent également un important système d'identifiants et de codes/métadonnées, permettant d'élaborer toute la structure du site, les liens ou renvois entre les différents contenus, la construction des recherches, les indexations, etc.

Ces données XML seront accompagnées de feuilles de transformation **XSLT** pour les mises en pages HTML, ainsi que de feuilles de styles CSS pour l'affichage écran.

Toutes les éditions du DAF, ainsi que les notices, seront structurées selon le même schéma XML.

Les autres types de données, notamment les nomenclatures et la base de formes fléchies, seront fournies sous forme de **tables** extraites de bases de données SGBD classiques (fichiers texte avec délimiteurs).

Toutes ces données étant facilement manipulables<sup>13</sup>, la construction d'un format de données « pivot », (d'échange), facilitant au mieux les développements et les réindexations automatiques, sera élaboré en liaison avec les besoins et outils du prestataire.

#### II. Consultation du site et fonctionnalités

#### II.1. Interface

L'interface du site devra être développée par l'intermédiaire du prestataire, mais en liaison avec l'Académie, et avec l'aide d'un **graphiste web** (interne à l'équipe ou en prestation externe), spécialisé dans la conception des interfaces web. Un cahier des charges ou *brief* spécifique devra être élaboré pour cet intervenant, après une phase d'analyse et de validation du cahier des charges définitif, à faire en amont, et en concertation avec l'Académie.

Cette interface revêt une grande importance dans la réussite du projet, et devra répondre aux caractéristiques suivantes :

• être claire, épurée, facile d'utilisation et immédiatement accessible pour un vaste public, non spécialiste et non nécessairement francophone.

<sup>&</sup>lt;sup>10</sup> Base terminologique, publiée par la DGLFLF, et à laquelle collabore l'Académie. <a href="http://www.culture.fr/franceterme">http://www.culture.fr/franceterme</a>

<sup>&</sup>lt;sup>11</sup> Trésor de la langue française informatisé, disponible sur le site de l'ATILF. <a href="http://www.cnrtl.fr/definition/">http://www.cnrtl.fr/definition/</a>

<sup>&</sup>lt;sup>12</sup> Un inventaire des versions numérisées du DAF est disponible ici : http://ahp.li/0d0c41de1d6ed016f623.html

<sup>&</sup>lt;sup>13</sup> Du côté de l'Académie, toutes les données du projet sont gérées dans une base numérique globale. Pour le développement du site web, ces données seront **exportées** via des scripts automatisés, qui peuvent être ajustés selon les besoins. Voir aussi le point V.4.



- être élégante, à la fois classique et moderne, et refléter l'origine institutionnelle du site, publié par l'Académie française.
- être conforme à **l'état de l'art** et des tendances récentes du webdesign, en termes graphiques comme en termes ergonomiques (fenêtres pop-in, scrolling, taille des typos, lisibilité, etc.), pour apparaître immédiatement comme tout à fait nouvelle.
- être conforme aux **usages** actuels de consultation des dictionnaires et ligne (boite de recherche, organisation de la page, liens de navigation, etc.), déjà bien en place (cf. exemples en annexe).
- être 100% compatible avec tous les principaux **navigateurs** du marché (Internet Explorer, Firefox, Google Chrome, Safari, etc.), dans leurs différentes versions en vigueur et sur toutes les plateformes (PC, Mac, Linux).
- être conçue en **design adaptatif** (responsive design), pour que le site puisse également être directement consultable, au moins sur les principales fonctions, depuis les navigateurs sur smartphones et tablettes (iOS, Android, etc.) et compatible SEO. Pour les smartphones, le site web devra autant que possible avoir l'apparence d'une **application mobile**<sup>14</sup> (voir aussi V.2.).
- donner le plus grand soin à la typographie et à la **présentation** du texte du *Dictionnaire*, dans un style à la fois élégant et classique (prestige, autorité et aspect historique du DAF), mais moderne et attrayant (percevoir immédiatement la modernité du design et de la présentation).
- anticiper dès le départ les **évolutions** et les versions futures du site, notamment l'intégration du corpus complet des 9 éditions du *Dictionnaire*, qui pose des problèmes spécifiques de consultation et d'organisation des données (au sein de la même interface, voir II.6)<sup>15</sup>.
- intégrer une **page d'accueil** (avec zone de recherche), donnant accès à différents types d'informations (voir ci-après)
- anticiper d'éventuelles évolutions, adaptations ou déclinaisons concernant l'accessibilité (voir II.2), ce qui peut conduire à prévoir une interface « bis » simplifiée.
- les propositions d'autres options graphiques ou ergonomiques, de la part du graphiste ou du prestataire, sont les bienvenues, et seront discutées en amont du projet, ainsi que leur faisabilité.

Plusieurs propositions d'interface et de design (3 ou plus) devront être fournies par le graphiste pour ce projet. Une réunion préparatoire sera réalisée très en amont du projet et dès son démarrage, afin d'analyser ensemble les principaux sites existants pris comme référence (voir en annexe) et les principes de ce qui est attendu.

En termes de remise du travail, l'interface devra être livrée en HTML/CSS (intégrant le design adaptatif), de manière à pouvoir être directement utilisable par les développeurs.

Une structure générale de l'interface est par ailleurs déjà pressentie (zone de recherche / zone d'affichage des articles / zone des blocs d'information...), afin d'intégrer les différentes fonctionnalités ; celle-ci est donnée en annexe, et pourra servir de base de travail.

#### Page d'accueil

Le design de la page d'accueil fera l'objet d'une attention spéciale, et devra être particulièrement lisible et clair. Elle devra être conforme aux usages actuels du web, proposer certains contenus éditorialisés, et devra comporter :

• la zone de recherche principale, permettant de rechercher immédiatement un mot ;

<sup>&</sup>lt;sup>14</sup> La réalisation d'une véritable application mobile, permettant la consultation sur smartphones du dictionnaire en ligne mais sans passer par le navigateur, sera sans doute envisagée ultérieurement, en reprenant les mêmes principes d'interface.

<sup>&</sup>lt;sup>15</sup> Rappelons que, dès la version 1 du site, celui-ci devra donner accès aux 2 dernières éditions (8<sup>e</sup> et 9<sup>e</sup>) du DAF.



- divers pavés d'informations autour du Dictionnaire, destinés à être réactualisés régulièrement, comme par exemple :
  - o une présentation du site ;
  - o les nouveautés et ajouts au Dictionnaire (fonctionnalités, mots nouveaux);
  - o les dernières rubriques *Dire, ne pas dire* parue ;
  - o un pavé dynamique sur les mots les plus consultés ;
  - des actualités particulières autour du *Dictionnaire* et de l'Académie, relayées depuis le site « corporate » ;
  - etc.
- d'autres types d'informations, comme par exemple l'aide en ligne, des liens vers les pages des réseaux sociaux, des liens vers d'autres ressources en ligne, etc.

Un travail spécifique de réflexion sur les éléments de cette page d'accueil devra être mené avec l'Académie, qui souhaite également recevoir quelques propositions ou modèles.

#### **Blocs d'informations**

Une des caractéristiques de la nouvelle interface sera de proposer, en partie droite de l'écran et en colonne, un ou plusieurs « blocs d'informations ». Ces blocs représentent des éléments d'informations complémentaires au Dictionnaire, c'est-à-dire des contenus annexes, internes ou externes au site. Ils peuvent être vus également comme des rubriques ou des fonctionnalités supplémentaires <sup>16</sup>.

Ces blocs peuvent varier selon les articles du dictionnaire et selon les informations disponibles ou non. Ils sont cliquables (soit en bloc, soit en intégrant des liens cliquables), pour pouvoir accéder explicitement et directement aux contenus ou aux fonctionnalités correspondantes.

Plus précisément, voici quelques-uns des blocs possibles :

- un bloc **plan de l'article** : pour les articles longs, ce bloc présente le plan (sommaire) des principaux sens de l'article affiché, avec les divisions sémantiques (I., II...., 1., 2...). Chaque division permet un accès direct au sens concerné, dans l'article.
- un bloc **conjugaison** : si l'article affiché est un verbe, le bloc donnera accès au tableau de conjugaison de celui-ci, en affichant par exemple les 3 premières formes (*je, tu, il...*) ;
- un bloc **notices** : il affiche, lorsque c'est possible, la liste des notices (voir plus haut) liées à l'article du dictionnaire consulté ;
- un bloc **orthographe**, donnant accès à des notices ou à des explications sur les Rectifications orthographiques, si le mot est concerné ;
- un bloc **liens externes** : il indique un certain nombre de liens externes (liens vers d'autres sites Internet) relatifs au mot concerné, qui peuvent être de toute nature<sup>17</sup> ; des sous-catégories pourront être insérées si besoin, par exemple **Terminologie**, **Francophonie**, **Synonymes**, etc. ;
- un bloc **éditions numérisées**, donnant accès à la page numérisée (fac-similé) contenant l'article consulté, pour les anciennes éditions (de 1 à 8) du DAF, disponibles par exemple (en mode image) sur Gallica.
- pour l'édition en cours du DAF (la 9<sup>e</sup> édition), ce bloc sera remplacé par l'accès à un **mode** « **page** », reproduisant à l'écran une mise en pages de type ouvrage imprimé (en 2 colonnes), contenant l'article en cours d'affichage<sup>18</sup>.

<sup>&</sup>lt;sup>16</sup> Cette structure est assez standard et a été développée dans plusieurs produits du Robert, notamment le *Petit Robert* version 4.

<sup>&</sup>lt;sup>17</sup> Par exemple, des liens externes seront proposés pour accéder directement à l'article en ligne du TLFi, à la banque de terminologie France Terme, au site de la BDLP (base francophone) etc., pour le mot concerné.



- un bloc historique, contenant l'historique des articles précédemment consultés.
- etc

Il s'agit d'une structure très souple et extensive, permettant les évolutions futures du site et qui permettra de proposer divers types d'extensions au DAF.

Toutes les informations nécessaires affichées dans ces blocs feront bien sûr partie des données fournies.

# II.2. Accessibilité

Une étude devra être menée, parallèlement à la conception de l'interface, concernant les questions d'accessibilité du site (possibilité de le consulter pour certains publics, notamment malvoyants), en reprenant les contraintes auxquelles le site devra répondre pour être conforme à l'état de l'art<sup>19</sup>.

Cette accessibilité concerne tous les aspects du site : éléments d'interface, lecture du texte à l'écran, navigation, liens hypertextes, etc. Il s'agit également, par exemple, d'être compatible avec les logiciels de synthèse vocale (lecture audio des pages web).

Cette compatibilité pourra prendre deux formes possibles : soit arriver à concevoir un site unique qui soit intégralement « accessible », soit développer une interface secondaire qui le soit, interface qui sera plus simple et sans doute nettement moins graphique, mais qui sera néanmoins fonctionnelle.

Selon les résultats de cette analyse, le critère d'accessibilité du site sera probablement être reporté à la version 2 ou 3. Il devra faire l'objet d'une option dans le devis de prestation.

# II.3. Mise en pages écran

La mise en pages des textes du Dictionnaire sera réalisée via des transformations classiques en XSLT, en générant un format HTML/CSS standard. Ces éléments de base seront fournis par l'Académie, et une maquette « statique » a déjà été réalisée en ce sens (voir VI.1).

Cette mise en pages s'appliquera sur tous les éléments textuels mis en œuvre dans le projet (voir plus haut), à savoir les articles des différentes éditions du Dictionnaire, les notices, etc.

En collaboration avec le prestataire, cette mise en pages HTML/CSS devra cependant être enrichie de manière à devenir « dynamique » et à intégrer certaines fonctionnalités, notamment :

- le codage des hyperliens et la navigation hypertexte générale (voir plus loin).
- l'ajout d'infobulles sur certaines informations.
- la navigation au sein d'un article, via le plan pour les articles longs (voir plus loin).
- les liens entre les articles des différentes éditions.
- les liens vers la conjugaison.

Ces informations supplémentaires seront codées dans les fichiers (XML et autres), et devront être réalisées au moment de l'indexation des données.

Les feuilles de styles CSS seront également optimisées de manière à assurer la compatibilité avec les navigateurs, une sécurisation contre le « piratage » (décodage) des données, ainsi que pour les formats d'impression et pour le design adaptatif.

Note : la mise en pages des **tableaux de conjugaison** devra également être prévue (voir II.5), ainsi qu'un format d'impression.

<sup>&</sup>lt;sup>18</sup> Ce mode « page », obtenu en réalité par une transformation XSLT particulière, est consultable sur la maquette déjà réalisée du DAF9, voir annexe VI.1.

<sup>&</sup>lt;sup>19</sup> On utilisera en particulier les fonctionnalités récentes disponibles en HTML5. Certains organismes proposent des aides techniques pour intégrer l'accessibilité.



#### Protection des données

Par ailleurs, une solution pourra être étudiée pour « protéger » les données de mécanismes de rétroingénierie, c'est-à-dire pour empêcher la reconstitution automatique de données structurées simplement à partir de l'analyse des pages HTML (analyse des styles, etc.).

Une protection contre le piratage global du site (aspiration des données) est également évoquée en IV.2.

# II.4. Recherches

Nous décrivons ici les différentes fonctionnalités liées à la recherche d'informations dans le site.

#### Recherche simple d'un mot

La recherche simple d'un mot dans l'une des éditions du DAF, qui bien sûr est la fonction première, s'effectuera le plus simplement possible, via une **zone de recherche** (zone de saisie) standard et selon les usages actuels du web (de type Google, Wikipédia, etc.). Plus précisément :

- un système **d'autocomplétion** sera mis en place, suggérant les mots les plus proches au fur et à mesure de la frappe ; les suggestions porteront en premier sur les « mots commençant par », mais d'autres possibilités seront à étudier<sup>20</sup>.
- cette recherche/autocomplétion devra neutraliser les accents, ponctuations et majuscules.
- la recherche s'effectuera dans une **nomenclature générale** (ou « super-nomenclature »), comprenant tous les mots de toutes les éditions du DAF ;
- un système sera élaboré pour permettre également la recherche d'un mot via l'une de ses **formes fléchies** (par ex. *eussent* → <u>avoir</u>, *heureuse* → <u>heureux</u>, *baux* → <u>bail</u>), notamment pour les verbes. Ce système s'appuiera sur la base générale des formes fléchies.
- par défaut, le choix d'un mot affichera en premier l'article de ce mot dans **l'édition la plus récente** du DAF, c'est-à-dire dans la 9<sup>e</sup> édition ou sinon la dernière édition où ce mot apparaît.
- un **correcteur d'orthographe** (même simple) sera intégré, permettant par exemple de récupérer des fautes de frappe.

Notons que, dans le fonctionnement de base, la saisie d'un mot (ou d'un début de mot) devra conduire directement au **positionnement** de la nomenclature générale au mot le plus proche, avec affichage automatique de l'article de dictionnaire correspondant.

On se reportera aux exemples de dictionnaires numériques pour analyser ces fonctionnalités dans les produits existants.

#### Recherche avec jokers

En utilisant la même zone de saisie, la possibilité d'effectuer une recherche avec « jokers » (caractères de remplacement) sera étudiée, selon une syntaxe du type :

- mots commençant par : par ex. anti\*
- mots finissant par : par ex. \*teur
- mots contenant, par ex. \*terr\*
- lettre manquante, par ex. ??a?t pour avoir les mots de 5 lettres avec un « a » en 3<sup>e</sup> position et un « t » final.

Avec cette syntaxe de recherche, le logiciel doit afficher une **liste de résultats**, extraite de la nomenclature générale ; ces recherches fonctionnent donc comme des **filtres**.

<sup>&</sup>lt;sup>20</sup> Dans certains dictionnaires numériques, le logiciel propose par exemple les **mots composés** contenant le début de mot qui a été saisi : par exemple en tapant *dieu* on obtient des composés comme <u>prie-Dieu</u>, <u>à Dieu vat</u>, etc.



Cette extension pourra être reportée en version 2 ou 3.

### Recherche dans d'autres index

Outre la nomenclature générale du DAF, d'autres index de recherche pourront être intégrés (soit en version 1 soit ultérieurement) au système de recherche simple, notamment :

- les nomenclatures des **notices** (Dire, ne pas dire, Questions de langues, voir plus haut);
- les index des **proverbes** ou des **locutions** du dictionnaire<sup>21</sup>.
- l'index des termes d la base France Terme.

Pour la recherche dans ces index, il devra être possible de taper un des mots pour accéder aux éléments, par exemple l'un des mots du titre des notices ou des locutions.

Les résultats de cette recherche seront organisés en **catégories**, de manière à séparer les nomenclatures du DAF de ces index supplémentaires.

#### Filtres par éditions

À partir de la version 2, avec l'intégration de toutes les éditions du DAF, il devra être proposé à l'utilisateur de pouvoir visualiser les **nomenclatures séparées** de chacune des 9 éditions, et de limiter ses recherches à l'une d'elles. Ce système fonctionnera comme un *filtre*, ou encore comme un système dit à facettes : en dessous de la zone de recherche principale, l'utilisateur pourra sélectionner l'édition choisie, pour ne visualiser ou rechercher que la nomenclature correspondante.

#### Recherche en texte intégral

Toujours en utilisant la même zone principale de saisie, un bouton supplémentaire pourra être disponible pour lancer une recherche du terme saisi **en texte intégral**, c'est-à-dire dans l'ensemble du texte des éditions disponibles du DAF. Plus précisément :

- par défaut, la recherche s'effectuera dans l'ensemble des textes de toutes les éditions du DAF intégrées au site ;
- la recherche pourra s'effectuer autant que possible de manière **lemmatisée**, c'est-à-dire qu'en cherchant par exemple le verbe <u>tenir</u> on obtiendra toutes les occurrences de toutes les formes conjuguées de ce verbe<sup>22</sup>.
- la recherche en texte intégral devra pouvoir fonctionner également sur **plusieurs mots**, ou bien sur une **séquence de mots** (en utilisant les guillemets), comme sur Google.
- si un **filtre** a été activé par l'utilisateur sur l'une des éditions (voir ci-dessus), la recherche ne portera que sur le texte intégral de cette édition.

Cette fonctionnalité pourra éventuellement être reportée en version 2 ou 3, selon les contraintes de planning.

#### Recherches avancées

Un système plus élaboré de recherches avancées (multicritères), portant sur le corpus des dictionnaires, pourra être envisagé en version 3. Outre la recherche en texte intégral, les critères de recherche pourront porter sur divers types d'informations disponibles sur les données, comme par exemple :

• les entrées et les mots de la nomenclature ;

<sup>&</sup>lt;sup>21</sup> Cette fonctionnalité est développée dans le *Petit Robert*. Elle permet par exemple, en tapant le mot « diable », de trouver toutes les locutions contenant ce mot, puis d'accéder directement aux articles du dictionnaire où elles apparaissent.

<sup>&</sup>lt;sup>22</sup> Le système de lemmatisation peut s'avérer délicat et incomplet sur les anciennes éditions du DAF, du fait des **graphies anciennes** des mots. Ce point sera étudié et amélioré ultérieurement.



- les catégories grammaticales ;
- le texte intégral, limité à certaines zones textuelles, comme par exemple l'étymologie ;
- des recherches par domaines (termes de mathématiques, de médecine, etc.);

Ce type de recherches a été développé notamment dans les produits du Robert (*Petit Robert* et *Grand Robert de la langue française*) et dans certaines autres versions numériques du DAF (voir annexes).

# II.5. Conjugaison

Le nouveau site du DAF proposera la **conjugaison intégrale** de tous les verbes du Dictionnaire — pour la 9<sup>e</sup> édition seulement (DAF9). Cette fonctionnalité, désormais classique sur le web, consiste à donner pour chaque verbe son tableau de conjugaison, à tous les modes et à tous les temps, et avec certaines options possibles.

Le réglage précis de ce module de conjugaison peut en réalité s'avérer délicat, du fait des multiples cas particuliers dans la conjugaison des verbes français. Il faut en effet tenir compte d'un ensemble de caractéristiques des verbes : choix de l'auxiliaire, catégorie grammaticale, participe passé invariable ou non, verbes défectifs, impersonnels ou pronominaux, élision et « h aspirés », etc.

En conséquence, en termes de données, la mise au point de ce module s'appuiera sur :

- la base de données de toutes les formes verbales de tous les verbes du DAF9 (formes données à la voix active).
- une table indiquant, pour chaque verbe, un ensemble de « marqueurs », chaque marqueur correspondant à un paramètre particulier et ayant un effet sur la génération du tableau de conjugaison.

Le développement consistera alors à :

- générer automatiquement les tableaux de conjugaison, de manière dynamique,
- les afficher à l'écran, selon une mise en pages prévue dans l'interface<sup>23</sup>,
- relier les tableaux de conjugaison aux articles du dictionnaire,
- proposer à l'utilisateur certaines options, permettant par exemple de choisir la **voix** (active, passive, pronominale, pour les verbes pour lesquels c'est possible) ou bien **l'auxiliaire** (pour les verbes admettant les 2 auxiliaires avoir et être).

Par ailleurs, une **maquette** a déjà été réalisée de ce module de conjugaison, maquette créée à partir d'un export XML de la base de formes fléchies et d'une transformation XSLT vers HTML/CSS. Cette maquette pourra être prise comme référence pour le développement final du module.

De nombreux sites web proposant ce type de module pourront également être pris comme référence.

# II.6. Navigation et liens

#### Accès aux différentes éditions

Le projet intégrant plusieurs éditions du Dictionnaire de l'Académie (2 éditions en version 1, et 9 éditions après), un système de **navigation entre les éditions** doit être mis en place. Ce système s'appuiera sur un système d'identifiants, qui reliera, pour un même « mot », tous les articles de ce mot dans les différentes éditions où ce mot apparaît.

Au niveau de l'interface, il devra donc être possible de passer d'une édition à l'autre sur un même mot. A priori, la solution retenue est celle qui avait été développée déjà par les éditions Redon (voir plus haut), via un système **d'onglets**:

<sup>&</sup>lt;sup>23</sup> En particulier, l'affichage des tableaux de conjugaison en design adaptatif devra être pris en compte.



9<sup>e</sup> 8<sup>e</sup> (1935) 7<sup>e</sup> (1878) 6<sup>e</sup> (1835) 5<sup>e</sup> (1798) 4<sup>e</sup> (1762) 3<sup>e</sup> (1740) 2<sup>e</sup> (1718) 1<sup>re</sup> (1694) ...

- comme indiqué plus haut, lors d'une recherche simple d'un mot, c'est l'édition la plus récente contenant ce mot qui est affichée en premier : la 9<sup>e</sup> édition dans le cas général, ou l'une des autres si le mot a disparu de la 9<sup>e</sup> édition. Selon le cas, l'onglet correspondant est donc activé.
- après cela, l'utilisateur a la possibilité de cliquer sur un autre onglet, pour afficher l'article du DAF dans l'édition correspondante.
- si le mot est absent de certaines éditions, les onglets correspondants sont visibles mais grisés/inactifs.

Notons que, en version 1, seuls les 2 premiers onglets, correspondant aux 9<sup>e</sup> et 8<sup>e</sup> éditions, seront proposés.

#### Navigation hypertexte

La fonction **d'hypertexte généralisé** sera mise en œuvre, qui apporte une valeur ajoutée considérable aux dictionnaires numériques. Ainsi, à l'intérieur du texte des articles affichés, il devra être possible de cliquer sur n'importe quel mot du texte pour naviguer dans le dictionnaire, c'est-à-dire pour accéder directement à l'article du mot qui a été sélectionné.

#### Plus précisément :

- cette navigation pourra fonctionner soit par **double-clic** (celui-ci ayant pour effet à la fois de sélectionner un mot puis de déclencher l'hypertexte), soit par des **liens hypertextes sur tous les mots** du texte (liens qui doivent être générés automatiquement à l'indexation des données).
- cette 2<sup>e</sup> solution, souvent mise en œuvre dans les dictionnaires numériques récents, a l'avantage d'être compatible avec les **interfaces tactiles**, et donc les smartphones et tablettes.
- l'hypertexte doit fonctionner avec **lemmatisation automatique**, c'est-à-dire qu'il doit être possible de cliquer sur n'importe quelle forme (pluriel, féminin, verbe conjugué) pour accéder au lemme (entrée) correspondant. Cette fonctionnalité peut être mise en œuvre en utilisant la **base de formes fléchies**, ainsi que la nomenclature générale.
- la recherche de l'entrée cible doit s'effectuer dans la même édition qui est en cours de sélection par l'utilisateur. Ainsi, si par exemple l'article en cours de consultation appartient à la 8<sup>e</sup> édition, l'entréecible doit être affichée également dans la 8<sup>e</sup> édition. Si jamais l'entrée-cible ne figure pas dans cette édition, le mot est alors recherché dans une édition plus récente, via la nomenclature générale.
- notons que l'hypertexte généralisé pourra être automatiquement désactivé (à l'indexation) sur certaines zones de texte, comme les transcriptions phonétiques, les mots en grec, etc.

#### Navigation entre le dictionnaire et les notices

Comme il est indiqué plus haut, le site intègrera un ensemble de notices, qui sont des développements autour des mots et de leurs usages (voir I.5). Au niveau des données, des liens hypertextes ont été constitués, permettant, depuis un article de dictionnaire, de renvoyer à une ou plusieurs notices et, inversement, à partir d'une notice de renvoyer vers des articles du dictionnaire.

Cette navigation s'effectuera dans le site de manière fluide, c'est-à-dire au sein de la même interface : les documents, qu'ils proviennent du dictionnaire ou des notices, s'afficheront en zone centrale, avec une mise en pages HTML/CSS. Bien entendu, les onglets donnant accès aux différentes éditions seront cependant masqués lors de la consultation des notices.

L'accès entre ces éléments s'effectuera comme suit :

• depuis un article du DAF, la liste des notices liées apparaîtra dans un bloc d'informations en colonne de droite (voir plus haut);



• depuis une notice, la liste des mots liés apparaîtra directement dans le texte, en fin d'article, avec des liens hypertextes.

# II.7. Fonctions bureautiques et de partage

#### Fonctions bureautiques

Au sein de l'interface, une petite barre d'outils proposera, comme il est très courant sur Internet, quelques fonctions bureautiques usuelles :

- imprimer l'article : cette fonction proposera un format d'impression optimisé, uniquement de l'article du dictionnaire consulté (ou du tableau de conjugaison, ou de la notice consultée, etc.), et avec une feuille de styles CSS adaptée ;
- augmenter ou réduire la taille des caractères : là encore, cette fonction sera optimisée par rapport au zoom standard du navigateur, en ne s'appliquant qu'au texte de l'article consulté ;
- aide : accès à une aide en ligne.
- historique : historique des articles consultés, ainsi que les fonctions suivant/précédent<sup>24</sup>.

D'autres fonctions pourront être proposées ultérieurement :

- favori : ajouter l'article consulté à une liste de favoris (voir ci-dessous) ;
- commentaire : laisser un commentaire sur l'article consulté.

# Fonctions de partage

Ces fonctions concernent essentiellement le partage sur les réseaux sociaux (Facebook, Twitter, etc.), ou bien le partage par courriel, et sont également très standards sur le web. Elles permettront à l'utilisateur de partager facilement l'article en cours de consultation, via un lien URL vers celui-ci (voir ci-après IV.2), comme par exemple pour un article Wikipédia ou de la presse.

# Comptes utilisateurs

Selon les possibilités techniques, on pourra envisager d'intégrer au site un système — facultatif — d'inscription des utilisateurs, via une création de compte simple (adresse e-mail / mot de passe). La création d'un compte permettra aux utilisateurs d'accéder à quelques fonctions supplémentaires, comme par exemple :

- la possibilité de poster des commentaires ;
- la possibilité de créer et de gérer des listes de mots ou de favoris, privées ou bien publiques. Voir par exemple ce qui est proposé sur le site Oxford (OALD) :

https://www.oxfordlearnersdictionaries.com/mywordlist/

Cette fonctionnalité n'a pas de critère d'urgence, et pourra être intégrée après la version 1.

# III. Méthodologie de travail

Sont indiquées ici quelques grandes lignes sur la méthodologie de construction et de gestion du site ; cet aspect peut bien sûr faire l'objet de propositions par le prestataire, selon ses outils et ses habitudes de travail.

<sup>&</sup>lt;sup>24</sup> Les fonctions suivant/précédent peuvent être mises en œuvre soit en utilisant les fonctions intégrées du navigateur (c'est-à-dire en utilisant les URLs), soit avec des fonctions et des boutons spécifiques au site.



# III.1. Indexation des données

L'Académie prend à sa charge la gestion des données, qu'elle livrera au prestataire via des exports automatisés depuis ses bases. Dans la phase initiale de développement du projet, des réunions préparatoires et des échanges réguliers devront permettre :

- de mettre au point des **formats d'échange** de ces données, entre l'Académie et l'équipe de développement; ces formats, structurés (de type XML, XSLT, CSS, HTML, bases de données, tables, etc.), seront décrits dans des documents de travail, explicitant la signification de la structure (balisage, identifiants, codes, champs, etc.) et de l'architecture des données (liens entre les données, appels croisés, etc.).
- une fois ces formats stabilisés, ils devront permettre de mettre au point des programmes automatisés d'indexation des données, permettant de générer automatiquement le site, avec son interface, et ce sans intervention aucune du prestataire sur les données elles-mêmes.
- ces programmes d'indexation devront pouvoir générer un listing d'anomalies (« log »), permettant de détecter des erreurs et de corriger les données (données, codes ou liens manquants, cohérence, complétude, etc.).

Il appartiendra au prestataire d'analyser en détail les données fournies par l'Académie, pour voir si ces données contiennent bien toute l'information suffisante pour réaliser les fonctionnalités voulues.

# III.2. Serveur de tests

Un serveur de tests, à accès sécurisé et privé, sera mis en place par le prestataire pour le projet. Ce site aura pour fonction de pouvoir tester le site complet avant sa publication en ligne. Il permettra notamment de tester la bonne intégration de l'interface, son fonctionnement sur différentes plateformes, les fonctionnalités, la bonne restitution des données et leurs contenus.

Ce site de tests devra pouvoir être regénéré automatiquement à partir de l'indexation ou de la réindexation des données, à tout moment.

# III.3. Mise à jour des données et réindexations

Bien que les données du *Dictionnaire* ne soient pas amenées à être mises à jour en tant que telles, les fichiers seront amenés à faire l'objet de livraisons successives :

- pour l'intégration progressive des derniers « cahiers » publiés par l'Académie sur la 9<sup>e</sup> édition (fin du dictionnaire), et ce jusqu'à fin 2019.
- pour des corrections diverses pouvant être effectuées, de contenu ou bien de balisage, notamment sur les anciennes éditions du Dictionnaire pour lesquelles des corrections seront inévitables.
- pour intégrer des enrichissements de balisages ou de nouvelles fonctionnalités.

Pour cela, les données devront pouvoir être livrées et réindexées automatiquement, à tout moment, afin de régénérer un site de tests, puis ensuite de publier une mise à jour du site public.

La réindexation des données sera effectuée par le prestataire. Si possible, un outil de type backoffice sera fourni à l'Académie afin qu'elle puisse elle-même déclencher ces réindexations, vérifier les serveur de tests et publier une mise à jour.

#### III.4. Backoffice

Il serait souhaitable que le prestataire puisse proposer un outil de backoffice, car celui-ci pourrait être avantageusement utilisé durant la phase de développement, puis ultérieurement pour la gestion du site.

Cette plateforme, qui serait au service des intervenants sur le projet (Académie, développeurs), serait utilisée pour :



- déposer les données (type FTP);
- lancer une réindexation des données, et consulter le fichier de « log » ;
- intervenir sur un certain nombre d'éléments du site, gérés par le prestataire et séparément des données fournies par l'Académie (par exemple des données de l'interface, ou des informations intégrées directement dans les pages HTML);
- en particulier, intervenir sur la page d'accueil du site, devant être « animée » par de nouvelles actualités, etc. ;
- consulter les statistiques de consultation sur le site (voir plus loin) ;
- mémoriser un ensemble de rapports de bugs ou corrections à effectuer ;
- consulter les commentaires utilisateurs et gérer ceux-ci (réponse, archivage, classement, etc.).

# IV. Hébergement et gestion du site

# IV.1. Hébergement

Le prestataire devra également proposer une solution d'hébergement du site. Celle-ci devra en particulier être capable de supporter des trafics potentiellement importants, et de piloter le site de manière à éviter tout problème de plantage, redémarrage, saturation en bande passante, etc.

Cette solution devra également proposer un système efficace de protection contre le piratage du site, du type « aspiration » des pages HTML par des robots, etc.

La solution proposée devra prendre la forme d'un forfait annuel pour le coût de l'hébergement, distinct du coût de développement du site lui-même.

# IV.2. Appels externes

#### Adresses publiques et URI

Un travail important devra être effectué sur la syntaxe des adresses URL/URI utilisées dans le site. Plus précisément :

- chaque article du dictionnaire, dans toutes les éditions, devra recevoir une URL unique, qui puisse être utilisée comme identifiant et comme référence à cet article, par exemple en Web sémantique<sup>25</sup> ou dans les résultats des moteurs de recherches/Google (permalink).
- la syntaxe de cette adresse pourra soit utiliser les identifiants internes existants déjà dans les données, soit une syntaxe utilisant le mot-entrée, du type :

http://www.dictionnaire-academie.fr/daf6/bonjour

http://www.dictionnaire-academie.fr/A6B1521 [où A6B1521 est l'identifiant interne de l'article « bonjour » dans le DAF6]

On prendra soin cependant de distinguer les homographies: mots strictement homographes (par ex. <u>poster</u>, v. et <u>poster</u>, n. m.) ou bien homographes via la neutralisation des accents et/ou des majuscules (par ex. <u>côte</u>, <u>cote</u>, <u>coté</u>, <u>côté</u>..., ou encore <u>b.-a.ba</u> et <u>baba</u>, etc.), via un système à trouver.

• ce système permettra également de pouvoir effectuer des appels externes au dictionnaire, de construire des applications diverses, des boîtes de recherche, des intégrations, etc. sur des portails externes au site ou encore des applications mobiles.

<sup>&</sup>lt;sup>25</sup> Notons que le Ministère de la culture a publié un « Guide des bonnes pratiques » à destination des éditeurs de contenus, afin d'indiquer une méthodologie de nommage des identifiants et de gestion des permaliens (<u>voir ici</u>).



- une syntaxe similaire devra être mise en place également sur le module de conjugaison, de manière à pouvoir appeler le tableau de conjugaison d'un verbe directement, par exemple :
  - http://www.dictionnaire-academie.fr/conj/epeler
  - ou http://www.conjugaison-academie.fr/epeler
- dans la mesure du possible, ces URL/URI seront également disponibles pour les autres contenus proposés dans le site, comme les notices.

Cependant, une réflexion doit être menée pour éviter que ce système d'URL/URI ne permette trop facilement le **piratage** du site, c'est-à-dire l' « aspiration » des contenus par un programme simple qui appellerait tous les articles successivement.

#### API

Le site devra également proposer une API, c'est-à-dire une boîte à outils permettant d'interroger les contenus de manière externe, en renvoyant par exemple un flux XML. Il s'agit là de technologies standards, mais qui sont importantes pour le projet.

Cette API pourra être relativement simple en version 1, en utilisant des protocoles standards, et devra être documentée, avec des exemples. On pourra s'inspirer de certaines API du même type, déjà proposées pour certains sites de dictionnaires.

#### Appel externe du dictionnaire

En particulier, ces outils devront permettre d'intégrer au site « corporate » de l'Académie française, sur la page d'accueil, une boîte de recherche centrale permettant d'effectuer directement une recherche dans le site du *Dictionnaire*. Il s'agira en réalité d'un appel externe au site, renvoyant à celui-ci, et utilisant une méthodologie d'hyperappel (ou de renvoi).

# IV.3. Statistiques d'utilisation

De manière assez classique, le site devra pouvoir être analysé au niveau des statistiques, et ce à deux niveaux :

- un niveau « externe », avec un outil de mesure du trafic de type Google Analytics ;
- un niveau « interne », permettant d'analyser les consultations des internautes sur le site, par exemple quels sont les mots consultés, les mots recherchés, le nombre d'utilisations de telle ou telle fonctionnalité, etc.

Sur ce second point, il serait souhaitable qu'un outil soit disponible via le backoffice, permettant de consulter ou d'extraire ce type de statistiques.

Egalement, certains éléments, comme par exemple les mots les plus consultés, pourront être utilisés pour animer la page d'accueil (voir II.1).

#### IV.4. Référencement et SEO

Le site, ainsi que la structuration des pages HTML de celui-ci, devront être optimisés pour un référencement optimum par les moteurs de recherche. Même si le but du projet n'est pas nécessairement de générer un maximum de trafic, il est important qu'il soit parfaitement référencé, facile à trouver, et qu'il « remonte » de manière optimale dans les moteurs de recherche, en fonction d'un certain nombre de mots-clés classiques (« dictionnaire », « définition », « académie », etc.). Une stratégie sur ces aspects devra être proposée par le prestataire.



# V. Autres aspects du projet

# V.1. Fonctionnalités additionnelles

En amont du projet, certaines fonctionnalités pourront être envisagées bien que non immédiatement développées, selon le savoir-faire et les propositions du prestataire. En effet, il pourra être utile d'anticiper dès le départ le développement futur de ces fonctionnalités, afin de faciliter ultérieurement ce développement.

Cela concerne par exemple:

- le développement d'une **interface multilingue**, notamment en anglais (ce qui, au niveau de l'interface, conduit à ne pas utiliser de boutons-images mais des textes);
- l'ajout de la **prononciation** audio des entrées ;
- l'intégration de la synthèse vocale (pour la lecture des articles et/ou pour l'accessibilité).
- des outils de **visualisation** des données (voir par exemple le site de *l'Oxford English Dictionary*).

#### V.2. Application mobile

Le développement d'une application mobile (iOS / Android) n'est pas incluse dans le présent cahier des charges ; en principe, l'utilisation d'un design responsif devrait procurer un confort de consultation suffisant pour les tablettes et smartphones, via les navigateurs intégrés.

Il est possible cependant que le développement d'une application soit envisagé ultérieurement, soit pour proposer une interface de consultation encore plus optimisée (et simplifiée) de la version en ligne, soit pour proposer une version consultable **off-line** (uniquement de la 9<sup>e</sup> édition).

Pour cette raison, le choix des technologies utilisées devra permettre d'envisager le plus simplement possible un « portage » ultérieur du site vers une application mobile, en réutilisant au maximum les données comme les développements. Le prestataire devra également proposer une solution pour le développement de cette application à moyen terme, si elle s'avère nécessaire.

# V.3. Liaison avec le site de l'Académie

Comme il est indiqué plus haut, le nouveau site de consultation du *Dictionnaire* sera un site autonome. Il devra cependant être lié au site officiel de l'Académie française (<u>www.academie-française.fr</u>), c'est-à-dire que le dictionnaire devra pouvoir être appelé directement depuis celui-ci, via une boite de recherche directement intégrée à la page d'accueil.

Cette liaison et cette intégration de la zone de recherche pourra être effectuée soit via un système d'hyperappel soit par un autre type d'API.

# V.4 Proposition d'un CMS

Pour le moment, l'ensemble des données du projet sont gérées du côté Académie par le rédacteur du présent cahier des charges, qui agit en tant que coordonnateur du projet et en assure toute l'informatique éditoriale. Il serait cependant important d'envisager, à moyen terme, la mise en place d'un CMS pour la gestion de ces données, puis le transfert de celles-ci dans le CMS, et enfin la livraison à l'Académie d'un outil permettant d'assurer la pérennité de cette gestion des données et la reprise ultérieure par un informaticien/chef de projet dédié.

Cet aspect est en dehors du présent cahier des charges, l'échéance étant plutôt au niveau de l'achèvement du *Dictionnaire*; cependant une réflexion et une étude parallèle pourront être menées en amont, si le prestataire dispose de solutions dans ce domaine.



#### VI. Annexes

#### VI.1. Maquette HTML

Une maquette de consultation du Dictionnaire est consultable à l'adresse suivante :

http://ahp.li/40b938fc5ec7cfddb80c.html

Cette maquette est « statique », c'est-à-dire qu'elle ne dispose pas de fonctions de recherche.

On peut cependant y trouver:

- la **nomenclature** complète du DAF9 (jusqu'à RIMBALDIEN), suivie de la nomenclature du DAF8 (pour la fin de l'aphabet).
- une maquette (provisoire) de mise en pages HTML du dictionnaire, obtenue par une transformation XSLT à partir des fichiers XML réels de la nouvelle base informatique, en construction.
- des liens hypertextes actifs, pour les renvois **internes** au dictionnaire (« Voir xxx »).
- des liens hypertextes pour des renvois **externes** au dictionnaire, à savoir les rubriques *Dire*, *ne pas dire* ainsi que *Questions de langue*, vers la base France Terme et également pour les mots concernés par les Rectifications de l'orthographe. Dans cette maquette, ces liens externes sont ajoutés à la fin des articles, mais il est prévu dans le nouveau site d'utiliser un système de blocs en colonne de droite (voir II.1.).
- pour les verbes du DAF9, les tableaux de **conjugaisons**, avec une mise en pages HTML (statique également) et des liens vers les séries verbales.
- l'index des **proverbes**, avec liens vers les articles concernés.
- l'accès à une 2<sup>e</sup> mise en pages du texte du *Dictionnaire*, reproduisant le format imprimé en colonnes (Journal officiel). Ce mode « page » pourra être utilisé comme équivalent à la consultation des facsimilés pour les anciennes éditions, en bloc de droite (cf. II.1.). Cette présentation est obtenue très simplement par une transformation XSLT à partir des fichiers XML.

# VI.2. Exemples de sites

On trouvera ci-après une liste (non limitative) de sites Internet de dictionnaires, pouvant servir de référence au projet de l'Académie française, que ce soit au niveau des fonctionnalités, de l'interface, de la mise en pages, etc. :

- les dictionnaires numériques des éditions <u>Le Robert</u>; ces sites sont à accès payant, mais proposent des versions d'essai gratuites; des versions de démonstration sont également disponibles, par exemple pour <u>Le Petit Robert</u>, <u>Le Grand Robert</u>.
- le site Larousse.fr.
- le site de l'OALD (Oxford Advanced Learners Dictionary) : <a href="http://www.oxfordlearnersdictionaries.com">http://www.oxfordlearnersdictionaries.com</a>
- le site de l'OED (Oxford English Dictionary) : www.oed.com. Exemple d'entrées : digital, mammoth.
- le site du Longman LDOCE : http://www.ldoceonline.com.
- Le site des éditions Collins: <a href="https://www.collinsdictionary.com/">https://www.collinsdictionary.com/</a>. NB: ce site propose une API.
- Le site du dictionnaire de la Real Academie Española : http://dle.rae.es
- Le site de Webster : <a href="https://www.merriam-webster.com">https://www.merriam-webster.com</a>
- Le site du <u>Multidictionnaire</u> (accès payant, <u>essai 30 jours gratuit</u>).
- Le dictionnaire québécois Usito (accès payant, essai 10 jours gratuit).
- Le portail du <u>CNRTL</u>, comprenant notamment le <u>TLFi</u> (Trésor de la Langue français informatisé).

#### Académie française



- Le projet Enccre sur l'*Encyclopédie* de Diderot et D'Alembert : <a href="http://enccre.academie-sciences.fr/ice">http://enccre.academie-sciences.fr/ice</a>
- Le portail de dictionnaires <u>ARTFL</u> élaboré par l'Université de Chicago, ainsi que leur projet <u>DVLF</u> (Dictionnaire vivant de la langue française).
- Le portail des éditions Garnier : <a href="https://www.classiques-garnier.com/numerique/">https://www.classiques-garnier.com/numerique/</a> (accès payant), qui comprend <a href="https://www.classiques-garnier.com/numerique/">1'ensemble des 8 éditions</a> du Dictionnaire de l'Académie.
- Le site de l'Encyclopedia Universalis : <a href="https://www.universalis.fr/">https://www.universalis.fr/</a>
- Le site (gratuit) des éditions Van Dale : <a href="http://www.vandale.nl/gratis-woordenboek/nederlands-frans/vertaling/woorden">http://www.vandale.nl/gratis-woordenboek/nederlands-frans/vertaling/woorden</a>

Ces sites sont de qualités variables, en termes d'interface et de fonctionnalités ; pour le présent projet, on se réfèrera plutôt aux versions numériques du Robert, de Collins et de Webster. Nous ne mentionnons pas ici les (très nombreux) autres sites de dictionnaires gratuits mais « anonymes » en ligne.

Notons que les sites gratuits en ligne sont pour la plupart extrêmement optimisés en matière de référencement.



## VI.3. Versions existantes du Dictionnaire de l'Académie

#### Portail CNRS

La version numérique « officielle » du DAF avait été confiée, dès 2000, au laboratoire ATILF du CNRS, à Nancy, car il publiait également la version informatisée du TLF *(Trésor de la langue française)*, qui venait d'être publiée en ligne. Le DAF a ainsi pu être publié avec le logiciel de consultation Stella, très performant mais qui n'a reçu malheureusement aucune évolution d'interface depuis près de 20 ans.

Il existe 2 éditions du DAF qui peuvent ainsi être toujours consultées dans cette interface :

- la 8<sup>e</sup> édition : <a href="http://atilf.atilf.fr/academie.htm">http://atilf.atilf.fr/academie.htm</a>
- la 9<sup>e</sup> édition (en cours) : <a href="http://atilf.atilf.fr/academie9.htm">http://atilf.atilf.fr/academie9.htm</a>

Par ailleurs, l'ATILF a développé depuis une seconde interface de consultation, plus conviviale, qui donne accès aux mêmes 8<sup>e</sup> et 9<sup>e</sup> éditions : <a href="http://www.cnrtl.fr/definition/academie9/">http://www.cnrtl.fr/definition/academie9/</a>, ainsi qu'à la 4<sup>e</sup> édition.

Cependant, cette seconde interface est également ancienne, très pauvre en fonctionnalités, et comporte de plus des erreurs matérielles sur les données, ainsi que des erreurs d'affichage.

C'est précisément pour remédier à ces 2 interfaces, qui ne donnent pas satisfaction, que l'Académie a souhaité effectuer une refonte de son interface de consultation, décrite dans le présent projet.

Notons que seul le portail du CNRS donne actuellement accès à la 9<sup>e</sup> édition, en cours.



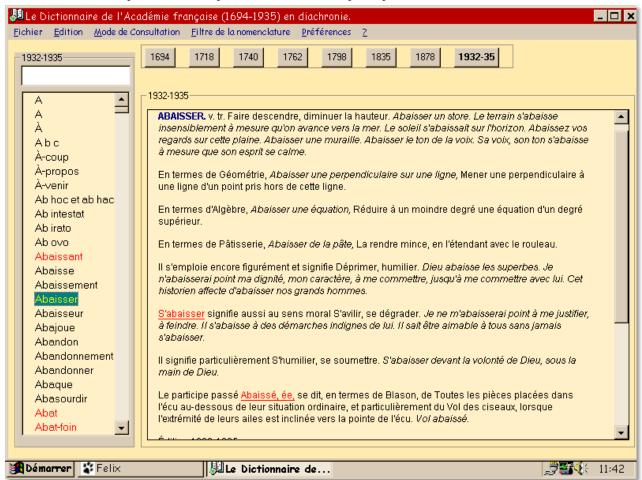
Interface de consultation du DAF9, via le portail CNRTL, comprenant 3 éditions du DAF



#### Logiciel Redon

Comme il est indiqué en préambule, les éditions Redon avaient publié à la fin des années 1990 un CD-ROM comprenant l'intégralité des 8 premières éditions du DAF. Ce logiciel, bien qu'ayant évidemment vieilli, proposait cependant des fonctions intéressantes de consultation et de recherches, dont certaines sont à reprendre pour le présent projet.

Des éléments plus précis pourront être fournis, ainsi qu'un exemplaire du logiciel qui est encore consultable. Nous donnons ici simplement une copie d'écran de l'écran principal :



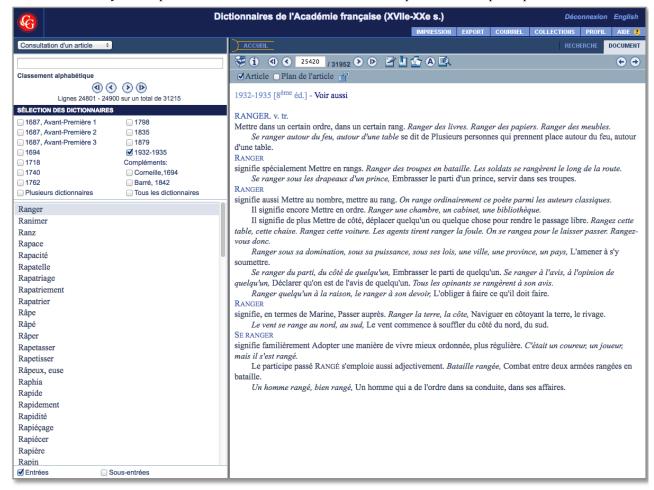
écran principal du CD-ROM Redon (1998), comprenant les 8 premières éditions du DAF



#### Site Classiques Garnier

Il existe un autre site de consultation des 8 premières éditions du DAF, publié par les éditions Classiques Garnier Numérique (voir ici). Ce site, à accès payant, est proposé sur abonnement aux universités et bibliothèques; il a été publié en 2007 et propose un certain nombre de fonctionnalités, dont certaines recoupent celles indiquées dans ce cahier des charges. Cependant, le site a une interface ancienne, n'est pas connu du grand public, et il manque certaines fonctionnalités tout à fait essentielles (comme la navigation hypertexte) qui ont été évoquées plus haut.

Nous ne décrivons pas ce site en détail ici, mais il pourra être utile, comme pour le CD-ROM Redon, d'en effectuer une analyse comparative. Nous donnons ci-dessous une copie de l'écran principal :



écran principal du site sur les 8 éditions du DAF, Classiques Garnier numérique

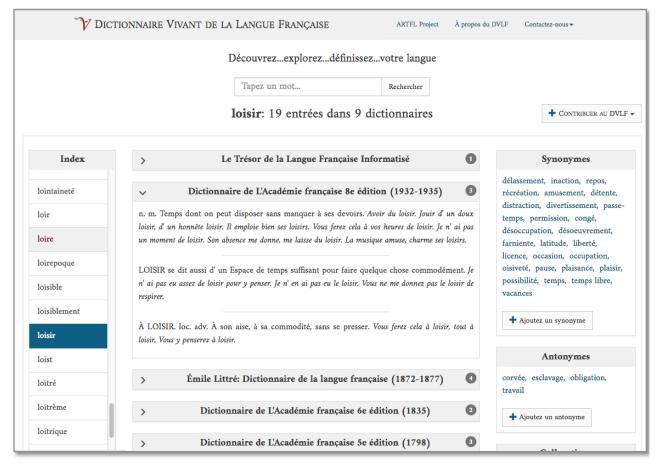


#### Site ARTFL

Il existe également un site proposant certaines des éditions du DAF, publié par l'Université de Chicago sous le titre « <u>Dictionnaires d'autrefois</u> ». Il contient les 1<sup>re</sup>, 4<sup>e</sup>, 5<sup>e</sup>, 6<sup>e</sup> et 8<sup>e</sup> éditions du DAF.

Ce projet, baptisé ARTFL (American and French Research on the Treasury of the French Language), est assez ancien (fin des années 1990), bien que le logiciel de consultation (PhiloLogic) ait été remis à jour depuis.

Ce corpus de 5 éditions du DAF est en fait consultable sous plusieurs interfaces, dans le cadre de ARTFL; on examinera par exemple l'interface du <u>DVLF</u> (*Dictionnaire vivant de la langue française*), qui regroupe un certain nombre de dictionnaires anciens sous une même interface.



écran principal du site DVLF, comprenant 5 éditions du DAF