

Balnéo : un projet de réseau informatique pour la veille néologique

Le contrôle de l'évolution des langues suppose des moyens d'observation rapides et interactifs dans le domaine clef qu'est la néologie scientifique et technique. La francophonie, qui a une longue expérience dans ce domaine, a donc souhaité se doter d'un réseau informatique international utilisant Internet pour la collecte et la diffusion des « attestations de néologismes ». Les principes d'organisation du réseau informatique *Balnéo* commandité par le Rint et réalisé par l'Université de Rennes II sont présentés à travers les différents acteurs qu'il fédère, la structure de l'information qu'il collecte (l'attestation) et les outils de saisie et de consultation qu'il utilise.

Termes-clés :
néologie scientifique et technique;
attestation de néologismes; Rint;
réseau informatique; Internet.

Les langues évoluent à la fin du XX^e siècle de façon prodigieuse, sous la poussée notamment des nouvelles techniques et de la révolution informationnelle en cours. Pour évoluer de façon harmonieuse, voire pour continuer d'exister, il leur faut s'informatiser et utiliser à leur profit les formidables outils de communication qui relient la planète.

Cette grande aventure des langues constitue un pari, mais aussi un risque.

Un pari car le véritable avenir des langues réside dans le plurilinguisme. Un risque car les maintenir au niveau des enjeux en cours exige de l'invention, des moyens, et une vision à long terme.

Il n'existait pas encore de moyen véritablement efficace pour observer les évolutions des langues dans le domaine de la néologie. Il est impossible à partir de là de pouvoir évaluer la vitalité réelle de leur lexique, et de répondre aux besoins de création de formes et de définitions. Une politique terminologique reste dans ce cas relativement aveugle, même si elle peut par ailleurs s'appuyer sur des indices ou des intuitions qui ne sont pas forcément à négliger.

Or, le politique, particulièrement le responsable d'une politique linguistique, doit nécessairement avoir une vision de la langue ou des langues à aménager, et définir des objectifs à atteindre. En ce qui concerne une langue comme le français, l'un des objectifs que se sont fixés les francophones est aujourd'hui d'observer l'évolution de ses

vocabulaires, particulièrement dans les domaines scientifiques et techniques. Sans vision d'ensemble de la néologie du français, en effet, il est vain de vouloir prévoir de façon rigoureuse toute évolution.

Disposer d'une base de données d'attestations de néologismes peut permettre en effet de faire des choix pertinents dans le cadre par exemple de la normalisation terminologique (fixation des termes), d'harmonisation terminologique (mise en correspondance de désignations autour de concepts), et plus généralement en matière de traduction de termes.

1 Les enjeux

La préoccupation de la francophonie a été depuis plusieurs années de résoudre la question de la néologie scientifique et technique. Pour faire évoluer la langue française, il est en effet nécessaire d'observer et d'analyser les néologismes qui apparaissent dans cette langue, soit par glissement de sens, soit par création de mots nouveaux.

À cette fin, des relevés d'attestations de néologismes, c'est-à-dire des formes repérées dans des contextes textuels (revues, journaux, etc.), ont été effectués par différentes équipes de recherche. L'une des expériences les plus représentatives est celle de la collection de fascicules *Néologie en marche* qui a lié pendant plus de dix années les équipes de l'Office de la langue française du Québec à celles du Haut Comité (puis du Commissariat général) de la langue française en France. Plus

récemment, le Centre de néologie et de terminologie du CNRS a élaboré, avec le soutien de la Délégation générale à la langue française, plusieurs «Cahiers de termes nouveaux», qui regroupent les relevés qu'il a lui-même réalisés. Dans la francophonie, des relevés sont également faits, notamment au Canada et au Québec.

Mais à l'évidence, ces expériences ne sont pas satisfaisantes, car elles demeurent trop ponctuelles, et ne parviennent pas à répondre aux besoins d'observation de la langue française aujourd'hui. En effet, il est indispensable d'avoir une vue large et précise de son évolution, que révèle la néologie, pour pouvoir intervenir à bon escient, cerner les priorités, et travailler de façon convergente avec les francophones.

Il faut pour cela globaliser le problème, travailler en temps réel et en mode interactif.

D'autant qu'il ne s'agit pas simplement de relever des néologismes français, mais aussi de faire un travail sur les néologismes de langue anglaise qui sont susceptibles de s'implanter en français. Il est également possible d'imaginer à terme que cette expérience puisse s'élargir à d'autres communautés linguistiques (hispanophonie, lusophonie, germanophonie), et aux langues partenaires de la francophonie.

2 Définition des objectifs

Afin de résoudre cette question du relevé des attestations de néologismes, il a été décidé par le Réseau international de néologie et de terminologie francophone (Rint), d'unir dans un premier temps les efforts de chacun des partenaires qui le composent en exploitant trois axes de travail :

- Localiser, dans chacun des pays francophones, les organismes et

personnes ayant par leur travail, à faire des relevés d'attestations de néologismes;

- Se répartir le travail de dépouillement en ciblant les domaines les plus porteurs de néologismes (techniques de pointe, informatique, communication, etc.);

- Élaborer un outil de travail qui puisse concentrer et diffuser l'ensemble des relevés effectués par les pays francophones.

C'est sur cette décision qu'a été lancé en 1994 le projet de base de données d'attestations *Balnéo*. La révolution technologique que constitue la constitution de réseaux mondiaux de communication de type Internet autorise aujourd'hui à penser l'organisation du travail en commun de façon radicalement différente. Dans ce cadre, il a été décidé de créer un système informatique appelée *Balnéo*, qui permette :

- D'établir un cadre de relevé commun;
- De stabiliser un format d'échanges de données commun;
- De diffuser sur Internet des fiches d'attestations de néologismes;
- D'en permettre la consultation;
- De les gérer dans une base de données;
- D'en effectuer la mise à jour;
- De travailler en temps réel ou en temps différé;
- De permettre un mode d'échanges d'information interactif.

Cet outil, conçu par l'Université de Rennes II, permettra de ce fait de recueillir les relevés d'attestations de néologismes effectués par les partenaires francophones du Rint qui ont accès au réseau Internet et de leur rediffuser l'ensemble des relevés réalisés.

Il sera ainsi possible à la fois d'alimenter et de puiser dans ce fonds d'attestations de néologismes. C'est donc ainsi une véritable banque interactive de données d'attestations de néologismes qui est proposée.

L'un des grands intérêts de cet outil est qu'il permet d'envisager le problème de la néologie du français de façon plus globale en répartissant les tâches de dépouillement entre plusieurs acteurs de la francophonie et au-delà. Son caractère interactif et partagé évite notamment que des domaines restent par inadvertance ignorés ou que deux partenaires dépouillent ou relèvent les mêmes choses. Il est possible d'envisager que puissent alimenter et avoir accès à cette base d'autres intervenants, comme la demande en est d'ores et déjà formulée (par le CNES par exemple). Cela aurait l'avantage de démultiplier les efforts des équipes du Rint, et d'enrichir ce fonds d'attestations néologiques dans une perspective encore inimaginable il y a quelques mois.

3 Principes d'organisation

3.1 Les principales fonctions

Le but du réseau *Balnéo* est de collecter rapidement et de diffuser largement au moyen d'un réseau informatique des attestations de néologismes lus ou entendus par des observateurs francophones. Dans la première phase de lancement il faut pour cela que :

- Le réseau informatique se calque sur la structure hiérarchique du réseau commanditaire (le Rint) qui est composé de modules nationaux qui travaillent avec des organismes dans lesquels des rédacteurs sont chargés de la production des «attestations»;
- La gestion du réseau informatisé soit automatisée au maximum : après son inscription (par une instance supérieure) sur le réseau, tout membre ajoute, consulte et commente librement les attestations sans intervention d'une tierce personne (il n'y a ni censeur ni «modérateur»);

- Des principes méthodologiques communs soient adoptés au sein du Rint pour la description d'un néologisme et que le réseau informatique collecte et fasse circuler les «fiches» ainsi définies (l'information échangée sur le réseau est purement textuelle, et le jeu de caractères utilisé est l'Iso-latin);
- Le réseau informatique garantisse que toute information diffusée en mentionne le rédacteur, le titulaire des droits de réutilisation, et informe sur les conditions de réexploitation et de cession;
- Seul le rédacteur d'une information puisse la modifier; mais une instance supérieure peut se substituer à lui en cas d'abus ou de défaillance;
- Tout membre du réseau puisse annoter une information produite par un autre membre. Les annotations sont jointes à l'information originale lors de sa diffusion. Elles ne peuvent être modifiées ou supprimées que par leur auteur.

3.2 Les usagers

- Les membres du réseau sont très dispersés (au moins en Europe, en Amérique et en Afrique) mais il sont supposés avoir un accès à Internet à partir de micro-ordinateurs sous *Windows*.
- Le travail de saisie et de mise à jour de l'information est long, il doit s'effectuer «hors ligne» avec exportation périodique des données vers la base centrale.
- La consultation de cette base et la rédaction d'annotations doivent pouvoir se faire «hors ligne» et «en ligne».
- Les rédacteurs sont essentiellement des utilisateurs de traitement de texte. Les outils locaux de saisie et les procédures d'import/export via Internet doivent donc être simples et conformes aux règles ergonomiques de *Windows* qu'ils connaissent déjà.

4 Le réseau informatique *Balnéo*

Développé dans le cadre très évolutif (voire mouvant) d'Internet, le système *Balnéo* a connu deux étapes bien distinctes. De juin à décembre 1994, un prototype a permis de tester la faisabilité d'un tel réseau. Il a été utilisé par les premiers modules nationaux qui avaient accès à Internet, notamment le module belge. Les observations et commentaires ont été utilisés pour concevoir le produit final qui devrait voir le jour pendant le premier semestre de 1996.

4.1 Principes de la première version (décembre 1994) : échanges en différé par messagerie

4.1.1 La production de l'information

La première version de *Balnéo* utilise uniquement une base de données locale (dite «base éditoriale») sous *Windows* pour la saisie des néologismes et la génération des messages administratifs. Cette base de données développée à partir d'*Access* de Microsoft permet la saisie des néologismes, leur recherche, leur mise à jour etc. et la formulation de requêtes.

4.1.2 L'envoi de l'information

Une fonction d'exportation exporte la base de données ou la requête dans un fichier ou dans le presse-papiers de *Windows*. Le résultat de cette exportation est transmis par un logiciel de messagerie (Mail, Eudora...) à une boîte aux lettres située sur un serveur central.

4.1.3 Le traitement des messages

Sur le serveur central un programme capte le message (selon le principe des «serveurs de liste» qui a inspiré cette première version). Il identifie l'émetteur et interprète le contenu du message. Si c'est une commande d'ajout, il enregistre les données jointes au message dans la base de données relationnelle du système central. S'il s'agit d'une requête (question), il interroge cette même base centrale et renvoie les réponses dans la boîte aux lettres de l'émetteur du message. Celui-ci lit son courrier, extrait la réponse dans un fichier ou le presse-papiers puis l'importe dans sa base éditoriale locale où il peut lire, annoter, éditer...

4.1.4 Avantages et inconvénients de cette méthode

Le procédé est simple à mettre en œuvre. Il se prête bien à la production des données (travail long sur la base locale puis envoi périodique en masse) mais il est lent et lourd pour les interrogations ponctuelles de la base centrale (pour vérifier par exemple si un néologisme a déjà été enregistré).

Le défaut majeur de la messagerie réside dans le procédé d'identification de l'émetteur : les logiciels de messagerie sous *Windows* ne contrôlent ni le nom ni l'«adresse Internet» de l'utilisateur, il est alors pratiquement impossible d'authentifier l'émetteur d'un message et donc de connaître son statut et ses droits. Sur ce point une expérimentation est en cours, qui permettra d'authentifier tout utilisateur entrant sur le réseau informatique (voir 4.2).

4.2 Les évolutions (mars 1996) : l'accès par le *World Wide Web* (W3)

Peu après la conception de cette première version, le standard HTML

2+ qui est à la base des services W3 (connus aussi sous le nom de «Web» ou, en français «cybertaine») a apporté un nouvel élément pour la réalisation de services interactifs : les formulaires. Grâce à eux, l'utilisateur d'un client W3 peut transmettre des informations à un serveur W3 où elles sont immédiatement traitées par un programme. Des développements informatiques spécifiques ont abouti par ailleurs à une solution satisfaisante dans le domaine de l'authentification de l'utilisateur (Le Meur 1995). Une nouvelle version de *Balnéo* plus interactive et plus «sécurisée» est donc en cours de développement.

Tout utilisateur reçoit lors de son inscription sur le réseau un mot de passe. Par la suite, quand il se connecte au moyen d'un client W3, il fournit ce code confidentiel. L'utilisateur authentifié peut procéder à toutes les opérations (administration, alimentation du réseau, annotation, interrogation) selon le statut et les privilèges qu'il a dans le réseau.

Quand il lit une fiche, il peut taper une annotation qui sera immédiatement ajoutée à cette fiche.

Un autre formulaire permet une saisie rapide et en ligne d'une attestation mais les limites de l'éditeur des clients W3 (*Netscape* ou *Mosaic*) rendent l'opération difficile pour certains types de champs complexes et répétitifs.

Autre évolution importante : la première version de *Balnéo* ne disposait que de la base éditoriale *Access* pour la saisie des attestations.

Les utilisateurs ont souhaité disposer d'outils plus proches de leur logiciel favori : le traitement de texte. L'exemple ci-dessous illustre la saisie d'un néologisme telle qu'elle pourra s'effectuer sous *Word 2* pour *Windows* :

Les données saisies ou importées dans la base éditoriale *Access* peuvent être exportées en format SGML, RTF ou Iso 2709 pour une réutilisation par une base de données documentaire, un

Microsoft Word - BALNEO2\ICHE.DOC

Fichier Edition Affiche Insère Format Outils Tableau Ecran ?

1 — DEBUT
TERME

2 — langue
gram
écrit-oral
*domaine
*domaine
attestation
source de l'attestation
année de l'attestation
situation
note sur la situation
*equiv<langue><note>
*note<source>

3 — FIN

moment-auteur
FRE
ncms
écrit
conception et réalisation
enseignement assisté par ordinateur
Le découpage d'un moment-auteur en modules obéit, certes à des nécessités informatiques, mais reste très proche d'une démarche qui tente de décrire, dans l'échange avec l'élève, une relation d'enseignement.
Revue du Palais de la Découverte. Sept conférences sur l'informatique. N° spécial, janvier 1985. Paris
1985
accepté
terme retenu par le Plan National EAO
pedagogical stage <langue> ENG <note> non attesté
L'auteur conçoit cette unité pédagogique en y intégrant les actions de l'élève et les éventuels résultats antérieurs. Le terme est utilisé dans le système auteur Diane.

Pg 1 Sec 1 1/1 A Li Co 5 100% NUM

1. Chaque élément d'information est constitué d'un paragraphe dont le style (*Terme* par exemple) indique la nature;
2. Un champ peut être répétitif : il y a ici deux domaines;
3. Un champ est parfois complexe : un équivalent est suivi de sa langue (obligatoire) et peut être suivi d'une note;
4. Cette icône lance un programme de vérification de la fiche en cours;
5. Cette icône vérifie toutes les fiches et les dépose dans le presse-papiers en vue d'une exportation;
6. Les icônes A et S Ajoutent ou Suppriment une fiche;
7. L'icône D permet de définir des «valeurs par défaut» afin d'éviter les saisies répétitives.

logiciel de PAO, une base diffusable sur CD-Rom, etc.

Sur le site central, la gestion des données administratives et néologiques est assurée par un SGBD relationnel (Oracle 7 en l'occurrence). Ce type de base est nécessaire pour la gestion courante (gestion des usagers, des codes confidentiels, des ajouts de fiches, des mises à jour, des suppressions, des annotations) mais il

présente des limites dans la consultation des néologismes (par exemple la recherche des composants d'un terme). Les néologismes seront donc transférés périodiquement dans une base WAIS (Wide Area Information Service). Un formulaire de recherche permettra de formuler des équations de recherche plus fines à partir du client W3.

4.3 Perspectives et échéances

Lors de son dernier Comité d'orientation en janvier 1996, le Rint a fait le point sur le développement de *Balnéo* et a pris plusieurs décisions qui engagent son avenir.

– La structure d'une attestation a été définitivement stabilisée. La liste des 17 éléments constitutifs a été fixée. Il ne reste qu'à établir des listes fermées de valeurs pour certains champs comme celui des catégories grammaticales.

– Le modèle d'une attestation formalisé en SGML va être mis à la disposition des modules et des organismes qui souhaitent alimenter *Balnéo* à partir de leur propre module de saisie.

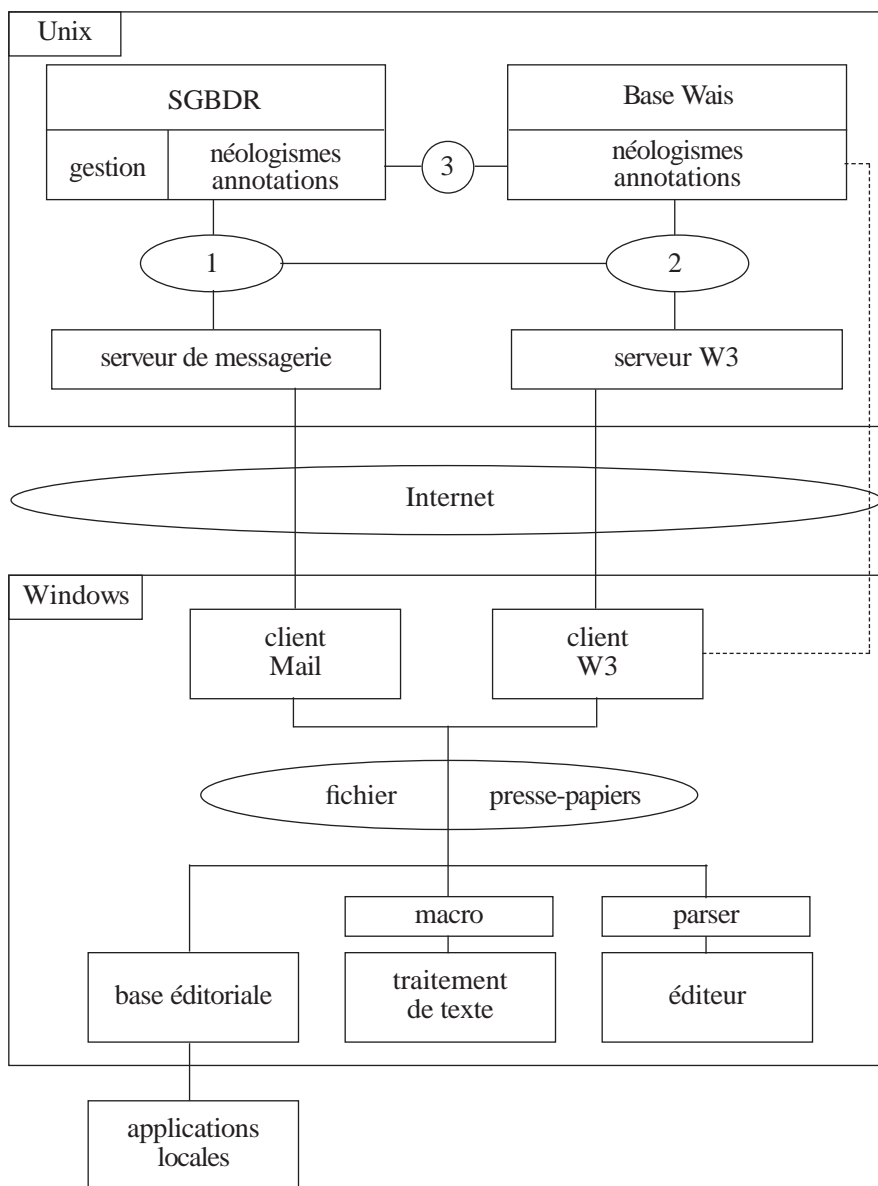
– Le système central, qui a été développé au Centre de ressources informatiques de l'Université de Rennes I, va être «porté» sur les machines de l'«exploitant» : le CRI de Rennes II (mars-avril 1996)

L'ensemble du dispositif devrait donc être opérationnel à la fin de ce premier semestre 1996. Il constituera un des éléments de la présence du Rint sur Internet.

Pour la première fois, il est possible de gérer selon un plan global les attestations de néologismes d'une langue, le français. Ce cadre sera étendu dans un premier temps aux langues partenaires de la francophonie, puis à d'autres langues.

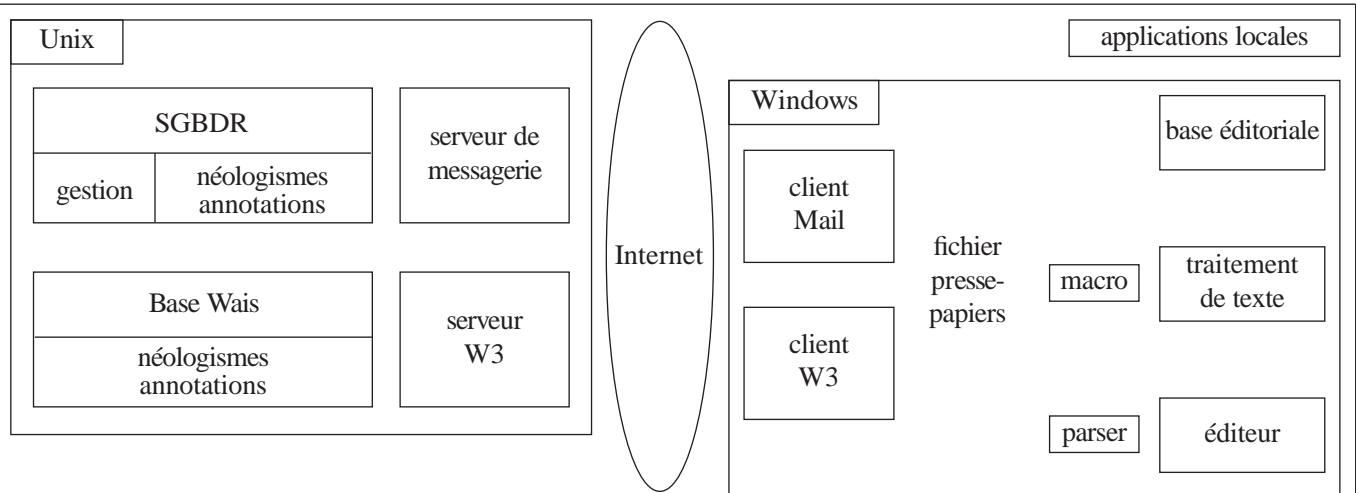
Parmi les autres développements envisagés figure une intégration plus poussée des relevés d'attestation de néologismes. Il devient envisageable en effet d'automatiser le repérage, le traitement, et la mise en forme de fiches d'attestation de néologismes à partir de banques de données textuelles, en plein essor aujourd'hui.

Balnéo : vue d'ensemble



Sur le site central trois applications ont été programmées :

- (1) reçoit, authentifie et traite les messages provenant de la boîte aux lettres ou du programme (2). Il met à jour la base de données relationnelle, renvoie un message à la boîte aux lettres du demandeur ou rend un message à (2).
- (2) authentifie un demandeur, transmet ses demandes à (1), rend les messages de retour au client W3.
- (3) convertit la base de néologismes au format WAIS et produit une base interrogeable par (2).



5 Quelques éléments de réflexion

Pour faciliter les développements informatiques et la maintenance du réseau, la structure de toutes les informations (administratives ou néologiques) est définie dans un modèle unique, une DTD SGML. Outre les avantages d'un formalisme unique, ce choix ouvre le réseau sur les outils standard développés autour de cette norme. On peut ainsi construire librement des éditeurs de messages *Balnéo* et utiliser les «parsers» pour vérifier leur conformité au modèle ou utiliser les langages de manipulation de données tels que *Balise* de la société AIS, *Omnimark* de Exoterica pour produire des dictionnaires, des CD-Rom, etc.

André Le Meur,
Université de Rennes II,
Loïc Depecker,
Délégation générale à la langue
française,
Paris.

Bibliographie

Le Meur (André), 1995 : «Méthodes et outils pour un travail de groupe sur Internet. Une application à la veille néologique», dans *Actes JRES95*, Comité réseau des universités, Unité réseau du CNRS, Chambéry novembre 1995 (à paraître).