

la recherche d'occurrences



La fonction «occurrences» est un développement de la fonction «rechercher» habituelle dans les traitements de texte. Il s'agit en quelque sorte, non seulement de rechercher, mais de **rapporter le contexte de ce qui est recherché**.

Idéographix offre de multiples possibilités de définir ce qui doit être recherché. Dans tous les cas, la séquence à rechercher est définie par l'utilisateur : lettre, groupe de lettres, mots, groupes de mots. Lorsqu'il s'agit de lettres, de groupes de lettres ou de mots, il est nécessaire de préciser si la séquence doit être retrouvée telle quelle ou localisée dans un élément plus vaste et dans ce cas

dans quelle position : «té» peut être un mot ou un groupe de lettres en début de mot (téléphone), en fin de mot (égalité), en milieu de mot (littérature) ; la recherche peut aussi ne pas se soucier de la position.

Dans tous les cas également, ce qui doit être rapporté est à définir par l'utilisateur. Il s'agit évidemment du **contexte** dans lequel l'élément recherché apparaît. Ce contexte peut être la phrase ou un certain nombre de mots avant et/ou après, incluant ou non un début ou une fin de phrase. On peut aussi intervenir sur l'affichage à l'écran et sur l'impression du résultat de la recherche, notamment en provoquant l'alignement vertical de l'élément recherché.

Idéographix permet encore de travailler sur les **cooccurrences** et non seulement sur les occurrences. L'intérêt est ici de rapporter un élément seulement si un autre élément existe dans son environnement, en spécifiant la nature de l'intervalle qui les sépare. On dispose au total d'un outil très puissant pour explorer un corpus de textes et en rapporter des événements linguistiques particuliers afin d'en faire une étude approfondie.

Enfin la recherche d'occurrences peut se lancer à partir d'une **liste** de mots, par exemple issue automatiquement d'une tablette.

Quelques utilisations

Les résultats d'une recherche d'occurrences sont affichés dans une page du traitement de texte et à leur tour utilisables comme un texte ordinaire. De ce fait, l'usage de base de la fonction consistera souvent à afficher l'ensemble des phrases d'une bibliothèque de textes où est employé un élément particulier (mot ou groupe de mots). On obtient ainsi un regroupement de phrases ayant en commun cet élément et qui permet d'en préciser le sens ou le rôle à partir du contexte. Donc un matériau facilement utilisable pour des explorations sur le fonctionnement de la langue écrite.

Mais il est tentant à ce stade de faire le dictionnaire de ces phrases ainsi réunies et donc de déterminer les mots les plus fréquemment associés à un autre. Si ce qui est rapporté n'est pas la phrase mais seulement les 3 ou 4 mots qui précèdent le mot recherché, il est facile de réfléchir ensemble sur le principe d'anticipation si important en lecture puisqu'un mot n'arrive jamais « comme ça ». La recherche peut également porter sur des éléments orthographiques au niveau du mot ou du groupe de mots. Par exemple, rechercher les mots qui précèdent ou qui suivent «les» pour découvrir quand il est déterminant et quand il est pronom.

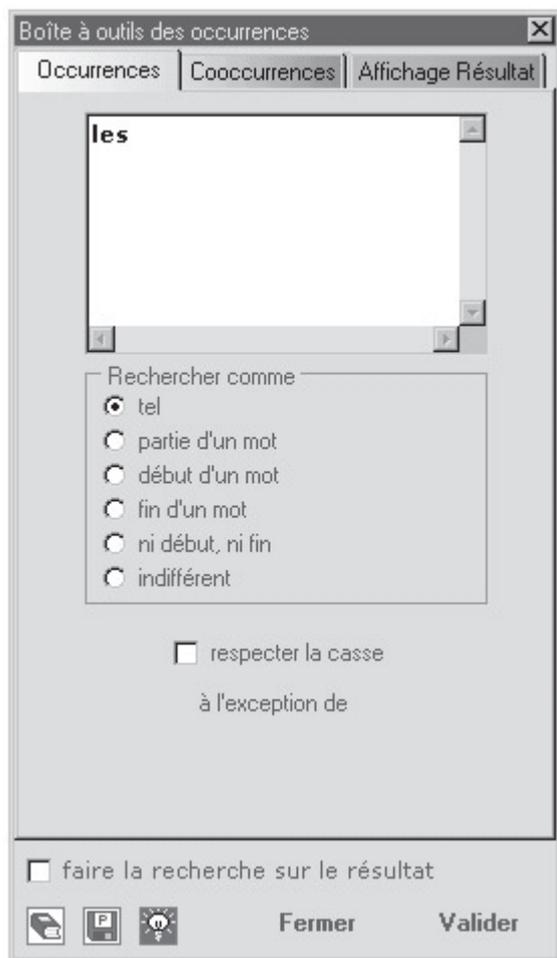


fig.1



fig.2

● Les occurrences Ou du particulier au général Du fortuit au construit Du contextualisé au décontextualisé

D'une part un texte (ou plusieurs), d'autre part une boîte à outils Occurrences.

(Fig 1) : l'onglet Occurrences est activé. Je propose une chaîne de caractères : **les**

Je choisis de le considérer comme un mot. La recherche portera sur tous les **les** ou **Les** (si on ne respecte pas la casse), déterminants ou pronoms, du texte.

Si **les** avait été coché ni *début*, ni *fin*, le résultat aurait pu faire apparaître *délester, molester...* Comme *début de mot* : *leste, lestes, lestemment, lester, lest...* Comme *fin de mot* : *fragiles, filles, difficiles, villes...* Comme *partie d'un mot* : tous les précédents. *Indifférent* : regroupe-

rait tous les cas précédents. À condition bien sûr que ces mots soient dans le (les) texte(s) considéré(s).

L'affichage (fig.2) fera apparaître le résultat de la recherche en gras, en rouge, dans des phrases.

Comme on est en traitement de texte, on pourra donc supprimer des phrases, ajouter un commentaire, imprimer, constituer un dictionnaire...

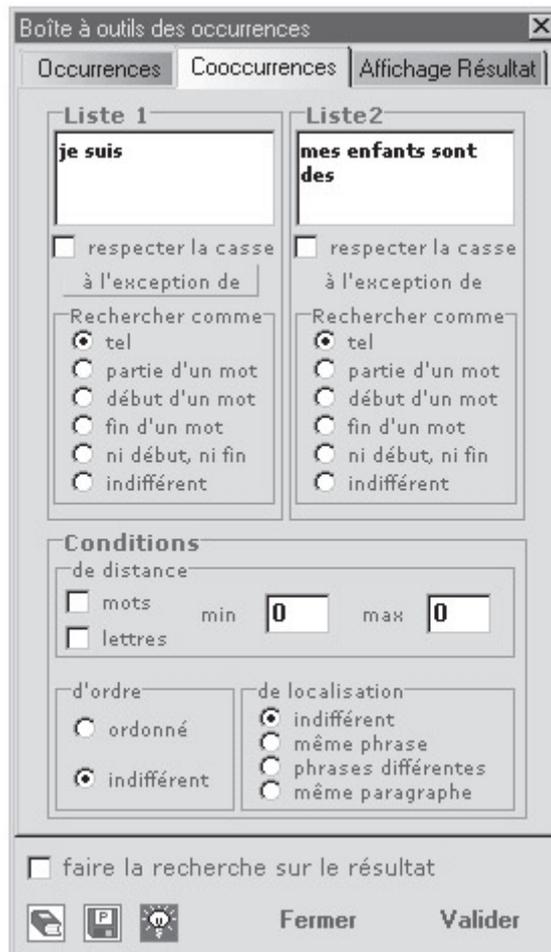
La recherche d'occurrences est d'abord un outil d'activité réflexive, de théorisation sur le fonctionnement de la langue écrite. Le premier intérêt est de recontextualiser un élément remarqué par exemple dans les dictionnaires et qui pose un problème de langue (ambiguïté sur la nature de cet élément, sur le sens, rôle des affixes, des flexions de conjugaison ou grammaticales...).

Si, par exemple, je cherche les occurrences de **a** et **à**, le traitement du résultat va conduire à une démarche

complètement opposée à celle du «Bled» pour traiter des homophones et homographes. Si je considère l'environnement, la place, le rôle dans la phrase, le sens, ces 2 mots vont apparaître comme des entités parfaitement distinctes, sans confusion possible puisque la référence linguistique porte sur l'écrit et non sur l'oral. Prenons un autre exemple, celui du **et**. Les occurrences sur les textes déjà rencontrés permettront de mettre en évidence que **et** joue bien un rôle de liaison, de coordination entre 2 mots de même nature et de même fonction, 2 propositions, 2 phrases, 2 paragraphes...

Chacun a pu constater qu'on peut aussi activer l'onglet cooccurrences. Il s'agit là de faire apparaître un élément, seulement s'il est en rapport avec un autre élément. L'exemple est pris dans *Poussin Noir* de Rascal (L'école des loisirs).

- **Je suis** un chien et **mes enfants sont** des chiots.
- **Je suis** une chatte et **mes enfants sont** des chatons.
- **Je suis** un porc et **mes enfants sont** des porcelets.



- **Je suis** un chèvre et **mes enfants sont** des chevreaux.

La juxtaposition des phrases met en évidence une structure, la chronologie des rencontres de Poussin Noir et les petits de chaque animal, ce dont a joué Rascal pour provoquer un effet sur le lecteur dans le texte.

Autre exemple dans le même texte : les cooccurrences de **Es-tu** et de **lui**, à propos du questionnement de *Poussin Noir*.

- **Es-tu** mon papa ? **lui** demande-t-il.
- **Es-tu** mon maman ? **lui** demande-t-il.
- **Es-tu** mon papa ? **lui** demande-t-il.
- **Es-tu** mon maman ? **lui** demande-t-il.
- **Es-tu** mon maman ? **lui** demande-t-il.

font apparaître l'alternance mâle / femelle et l'irrégularité pour le dernier animal (construction volontaire de l'auteur), les masculin/féminin sur les groupes nominaux, la structure identique, la pronominalisation et l'ambiguïté de genre pour tu et lui, l'inversion du sujet... Autant d'éléments qui pourront faire l'objet de remarques, discussions, vérifications dans d'autres textes, énonciations de règles selon le niveau du groupe concerné, ses connaissances préalables, la programmation de l'enseignant.

Ces quelques exemples donnent une idée de la puissance de l'outil et de son intérêt pédagogique pour le maître et les élèves. C'est un point de départ pour faire des constats sur l'écriture de l'auteur, la construction du texte, le système génératif de la langue, ses codes syntaxiques et orthographiques...

Lorsque, fortuitement parce que lié à ce texte précis, dans le dictionnaire des terminaisons, apparaissent les mots terminés par **nt**, la contextualisation par les occurrences permettra de classer ces mots en fonction de critères linguistiques. Même chose pour **ons**.

piaulant	1
marrant	1
Devant	1
Cent	1
répondent	1
découpent	1
ont	2
seront	1
front	1
sont	12

- *attendons*

Viens vite, mon poussin, nous t'**attendons** !

- *illustrations*

illustrations de Peter Elliot

- *chatons*

Je suis une chatte et mes enfants sont des **chatons**.

- *canetons*

Je suis un cane et mes enfants seront des **canetons**.

les mots de la famille de lait peuvent se trouver aussi à partir de la racine **lact**.

La *Recherche d'occurrences* est donc bien dans la cohérence d'un **bureau de lecture** et de **traitements** du texte aux côtés des *dictionnaires* et de la *Recherche lexicologique*.

André MOUREY, Gilles MONDÉMÉ

Si je reprends le mot **sont** qui apparaît 12 fois, la contextualisation fait apparaître les différents sujets : pronom, groupe nominal au pluriel, les formes affirmatives et négatives,...

- Cent œufs ont éclos dans la couveuse des fermiers Vitellus.

- 99 poussins **sont** nés jaunes.

- Ils **sont** jaunes et moi, je suis noir.

- Ce ne **sont** pas mes parents !

- Je suis un chien et mes enfants **sont** des chiots.

- Va donc voir dans l'étable si tes parents n'y **sont** pas, suggère Kaboul.

- Je suis une chatte et mes enfants **sont** des chatons.

- Va donc voir jusqu'au verger si tes parents n'y **sont** pas, suggère Agrippine.

- Je suis un porc et mes enfants **sont** des porcelets.

- Va donc voir près de l'abreuvoir si tes parents n'y **sont** pas, suggère Jacques-Henri.

- Je suis un chèvre et mes enfants **sont** des chevreaux.

- Va donc voir à la rivière si tes parents n'y **sont** pas, suggère Johanna.

- Va donc voir dans le bois si tes parents n'y **sont** pas, suggère Calypso.

On voit aussi qu'il y a un intérêt certain à élargir la recherche aux autres textes rencontrés pour avoir un début de généralisation statistique et mettre en évidence des régularités, des exceptions, et ainsi permettre à l'élève d'augmenter son répertoire graphique du français. Ce sont ces manipulations du particulier qui lui offrent la possibilité d'aborder le code graphique dans sa plus grande décontextualisation, en particulier avec un autre outil proposé par *Idéographix* : *la recherche lexicologique*. Ainsi **lait** rencontré dans *Poussin Noir*, recherché comme groupe de lettres, en position indifférente, dans les catégories Noms et Adjectifs sera représenté dans 20 formes différentes (32 avec les pluriels), toutes de la famille de lait mais avec une exception : laiton. Il faudra attendre une autre occasion ou apporter cette information,