

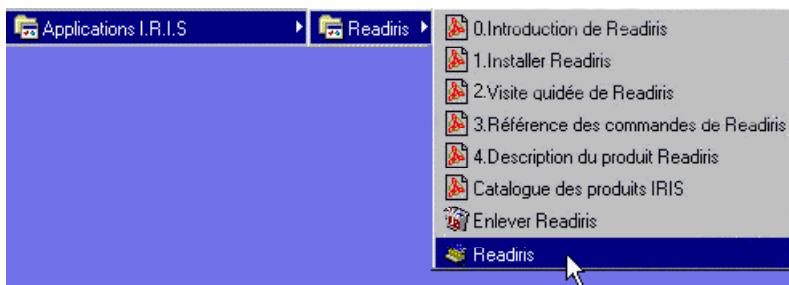
# Chapitre 2

## VISITE GUIDÉE

Readiris est un produit sophistiqué de reconnaissance optique de caractères (OCR) doté de nombreuses caractéristiques avancées. Nous envisagerons dans ce chapitre les principales applications de Readiris ainsi que de nombreuses astuces pour mieux l'utiliser.

### DÉMARRAGE DU LOGICIEL

Cliquez sur l'application Readiris dans le sous-menu "Applications I.R.I.S. - Readiris", ou cliquez sur le raccourci vers Readiris sur votre bureau Windows.



L'écran de démarrage ainsi que la fenêtre de l'application Readiris apparaissent. L'écran de démarrage affiche la version du logiciel ainsi que des messages liés aux droits d'auteur. Cliquez avec la souris n'importe où pour faire disparaître cet écran.



La fenêtre suivante concerne l'assistant d'OCR. Cliquez sur Annuler pour le moment.

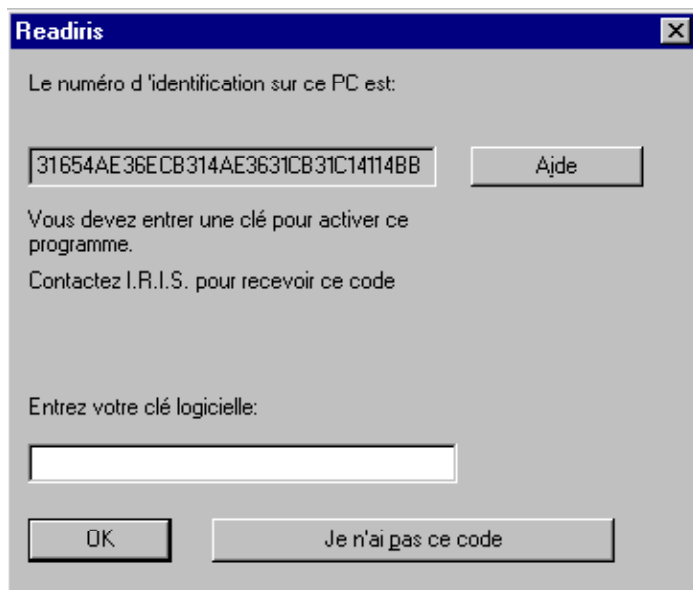
## LE PREMIER DÉMARRAGE

---

En fonction de la version acquise du logiciel, le premier démarrage peut être particulier: on peut vous demander d'enregistrer votre licence.

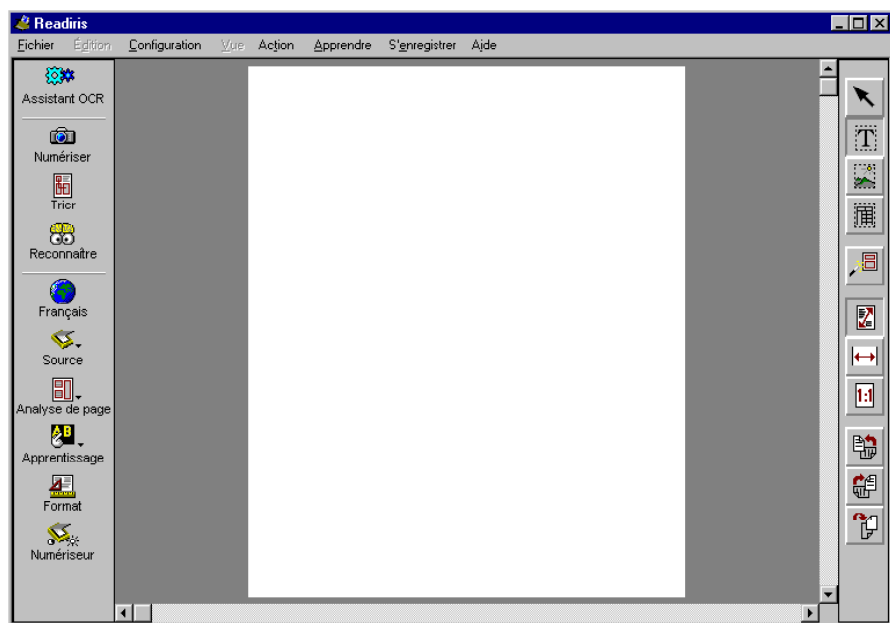
Si c'est le cas, l'utilisation de Readiris est limitée à un mois, et en vous enregistrant, I.R.I.S. vous enverra gratuitement une **clé logicielle** afin de pouvoir continuer à utiliser le logiciel après 30 jours.

Votre **numéro d'identification** est nécessaire à la création de la clé logicielle. Assurez-vous que ce numéro est disponible ou mentionné quand vous enregistrez votre licence.



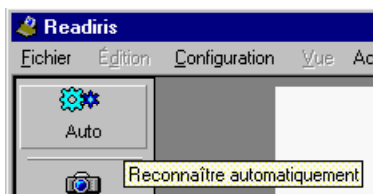
## DÉCOUVREZ L'INTERFACE DE READIRIS

La fenêtre d'application de Readiris ne contient pas seulement les **menus de commande** mais aussi deux barres de boutons qui donnent accès aux commandes fréquemment utilisées. Initialement, certains menus de commandes sont grisés: ils concernent la prévisualisation. Aussi longtemps qu'aucune image n'est ouverte, ils ne sont pas accessibles.



Il en est de même pour la **barre d'outils image** à la droite de la fenêtre d'application: elle contient toutes les commandes utiles à la prévisualisation. La **barre d'outils principale** sur la gauche de l'écran permet un accès rapide à toutes les commandes générales fréquemment utilisées.

Pour savoir quelle commande correspond à quel bouton, tenez le pointeur de votre souris sur lui pendant un moment: la **bulle explicative** de l'outil s'affichera.



Enfin, la zone centrale de la fenêtre ou la **zone d'image** est l'endroit où les images numérisées sont affichées. Vous pouvez glisser-déposer des fichiers image dans la zone image pour les reconnaître.

## DÉBUTER AVEC UNE PREMIÈRE LEÇON

La meilleure façon de se familiariser avec le logiciel Readiris est indubitablement de l'utiliser. Un certain nombre d'**images prénumérisées** est fourni avec le programme; il vous permettent de commencer son utilisation même s'il n'y a pas de scanner connecté à votre ordinateur. Voyons ce qu'il en est.

Le bouton "Source" sur la barre des boutons détermine si vous allez utiliser un scanner ou une image prénumérisée comme source image. Readiris donne la possibilité d'ouvrir des images TIFF (non-comprimées, comprimées en packbits, Groupe 3 et Groupe 4), des images Paintbrush (PCX) et des fichiers Windows bitmaps (BMP). Seules les images en noir et blanc sont acceptées. Cette possibilité est particulièrement utile pour convertir vos **télécopies** en fichiers texte éditables.

Comme nous allons ouvrir une image préscannée, vous devez sélectionner la disquette et non le numériseur comme source d'image.



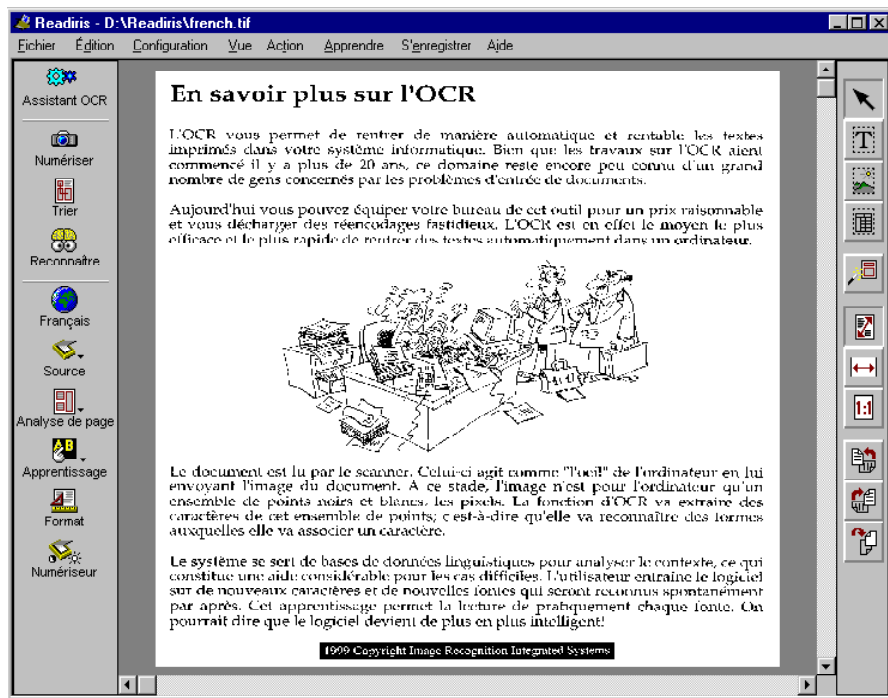
Ensuite, cliquez sur le bouton "Ouvrir". (Quand vous sélectionnez la disquette comme source d'images, le bouton "Numériser" est remplacé par le bouton "Ouvrir" et la commande correspondante "Numériser" sous le menu "Action" est remplacé par la commande "Ouvrir".)



Vous pourriez également sélectionner la commande "Ouvrir" du menu "Fichier" et ouvrir directement une image préscannée – ceci fonctionne même lorsque votre numériseur est utilisé comme source d'image.



Vous êtes invité à sélectionner un fichier image. Sélectionnez le fichier FRENCH.TIF dans le dossier Readiris. L'image est lue depuis le disque et progressivement affichée dans la zone image.



Une troisième façon d'ouvrir des images prénumérisées est d'utiliser la méthode du **"glisser-déposer"**. Faites glisser une image depuis l'Explorateur Windows en maintenant le bouton gauche de la souris enfoncé et amenez-la dans la zone image de Readiris pour l'ouvrir directement.

La barre d'outils image à la droite de la fenêtre d'application de Readiris contient toutes les commandes dont vous avez besoin pour la prévisualisation: les outils pour indiquer les zones d'intérêt, pour faire pivoter l'image, pour agrandir l'image, etc.

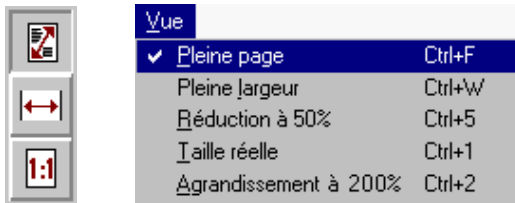


## AGRANDIR L'IMAGE

---

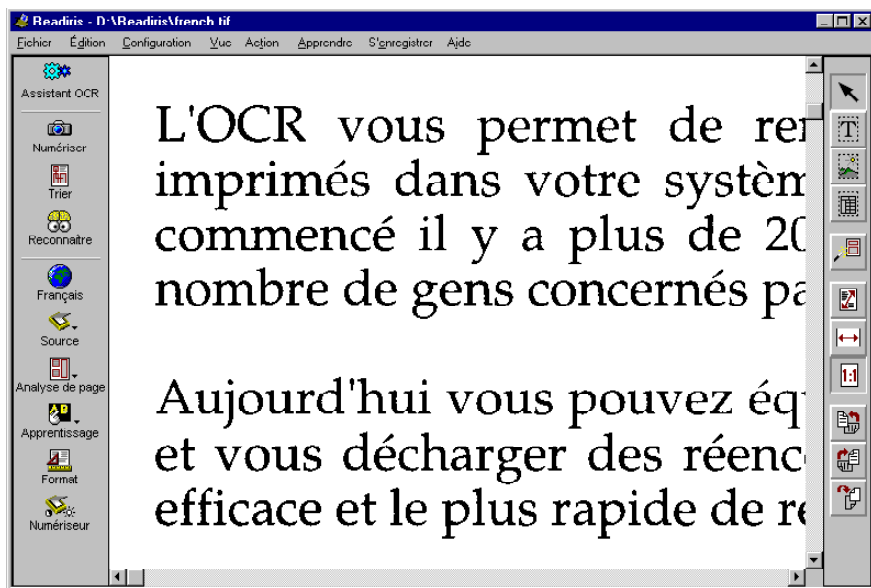
Readiris a plusieurs commandes qui vous permettent d'agrandir l'image numérisée, par exemple pour vérifier la qualité de numérisation.

La barre d'outils image contient des boutons qui permettent d'effectuer un zoom en taille réelle, d'ajuster l'image à la largeur de page et d'ajuster l'image entière dans la fenêtre d'aperçu. Le menu "Vue" contient les mêmes commandes et ajoute deux niveaux de zoom supplémentaires: vous pouvez afficher l'image à 50% et 200% de sa taille actuelle. A sa taille actuelle, un point écran correspond à un point image. (Des raccourcis existent pour tous les niveaux de zoom!)



Enfin, vous pouvez en cliquant avec le bouton *droit* de la souris sur une région de l'image numérisée en faire directement un agrandissement à sa taille réelle. En cliquant une seconde fois sur le bouton droit, vous annulez l'opération que vous venez d'effectuer.

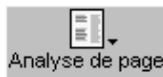




## UN, DÉCOMPOSITION D'UNE IMAGE NUMÉRISÉE

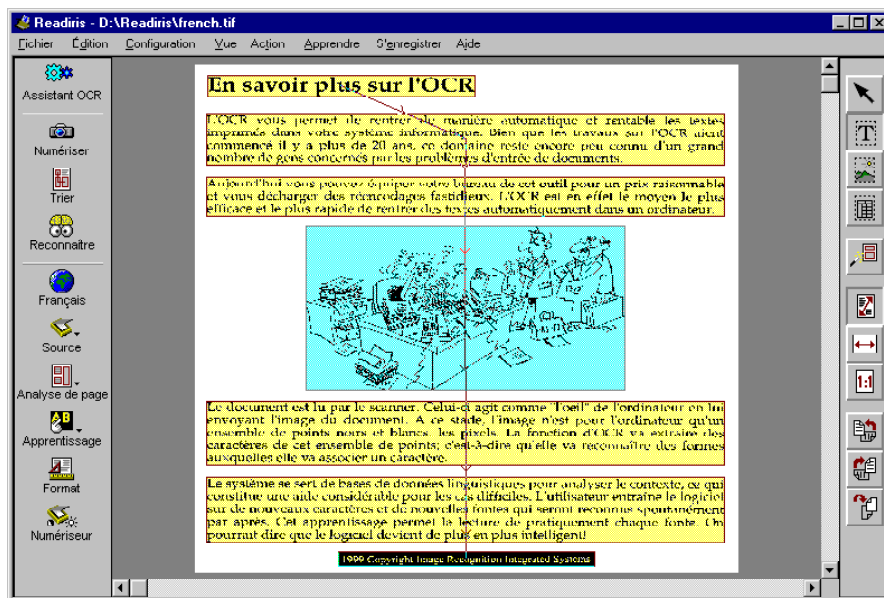
Maintenant que l'image est numérisée, vous pouvez définir des parties de textes que vous souhaitez lire en dessinant des "fenêtres", autour des zones d'intérêt.

En fait, Readiris effectuera automatiquement cette opération quand l'option "Analyse de Page" est activée sur la barre d'outils principale.





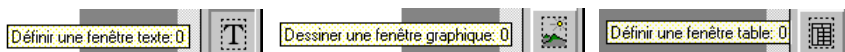
La décomposition automatique de la page est particulièrement utile lorsque des **textes sous formes de colonnes** ou à la mise en page complexe - qui pourrait inclure des graphiques et des tableaux - sont reconnus.



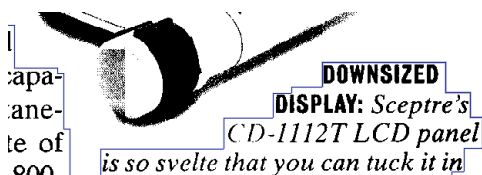
L'analyse de page utilise trois **types de fenêtres**: les fenêtres texte, les fenêtres graphiques et les fenêtres tableaux. Readiris distingue les blocs de texte, les tableaux et les zones graphiques qui contiennent des photos, des illustrations, etc. (La sauvegarde des graphiques et la reconnaissance des tableaux seront expliqués plus tard.)

Un **code couleur** indique le type de fenêtres: les fenêtres textes sont jaunes, les fenêtres graphiques sont bleues et les fenêtres tableaux sont pourpres.

Le nombre de fenêtres est indiqué à tout moment dans les bulles explicatives des outils "Fenêtre Texte" et "Fenêtre Graphique".



L'analyse de page est rapide et hautement précise, elle accepte que les lignes ne soient pas tout à fait rectilignes, et elle trace des figures complexes, "irrégulières".



L'analyse de page détectera même des zones où vous aurez un **texte blanc sur fond noir**. La reconnaissance de ce type d'insertions ne pose pas de problème: l'image apparaît à l'écran telle qu'elle est numérisée, mais Readiris "inverse" l'image pendant la phase de reconnaissance pour reconnaître de tels blocs de texte. (Vous pouvez grâce à votre numériseur générer des images complètement inversées pour les pages ayant du texte blanc sur fond noir. Voir plus bas.)

## UN ET DEMI, TRIER DES FENÊTRES

Readiris ne détecte pas seulement les différents blocs, mais en plus le logiciel les trie: les blocs sont triés du haut vers le bas, de gauche à droite afin de bien traiter les documents multicolonnés.

Vous pouvez évidemment modifier l'ordre de **tri**. Pour ce faire, cliquez sur le bouton "Trier" sur la barre d'outils principale. Le curseur prend la forme d'une main qui pointe dès que le "mode de tri" est activé.



Cliquez à l'intérieur de la fenêtre pour les inclure. Les fenêtres sur lesquelles vous ne cliquez pas sont tout simplement ignorées, exclues de la "lecture". La différence entre les fenêtres sélectionnées et non sélectionnées est claire: les

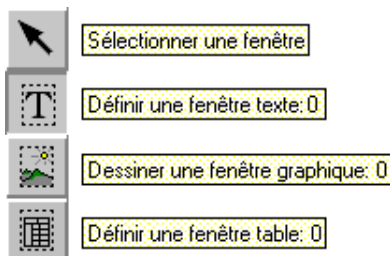


(Sélectionnez la langue du document *avant* d'effectuer l'analyse de page quand vous traitez des documents asiatiques. Des routines spécifiques sont utilisées

pour ces langues: l'espace interligne des documents asiatiques est généralement plus grand que dans les documents occidentaux, le texte est composé de petites icônes ("idéogrammes") qui pourraient facilement être vues comme des zones graphiques dans les documents occidentaux, et le texte peut se lire de haut en bas, de droite à gauche.)

## DEUX, FENÊTRAGE MANUEL D'UNE IMAGE

L'analyse de page traite automatiquement le fenêtrage d'une page numérisée. Autrement, vous pouvez dessiner manuellement des zones avec les **outils de fenêtrage** de Readiris.



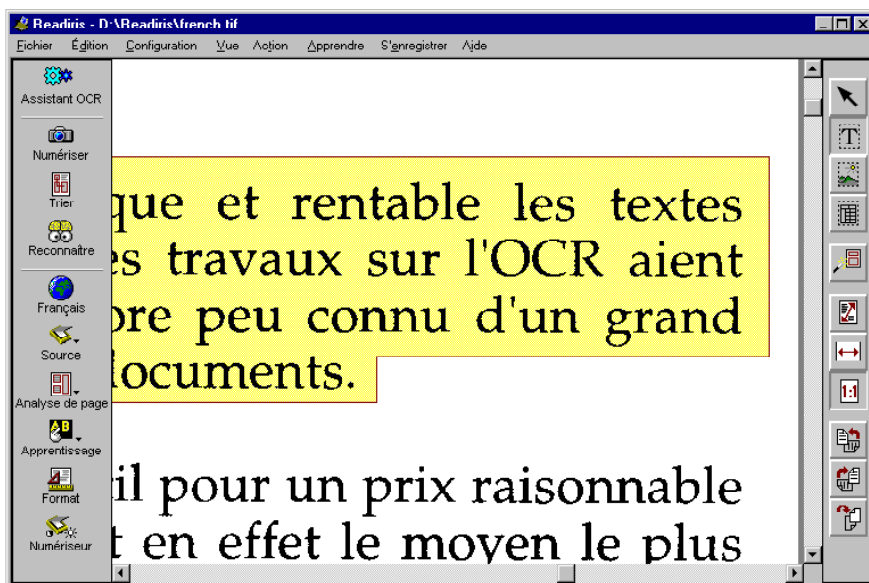
Pour **créer** un rectangle autour d'un bloc à reprendre, sélectionnez l'outil correspondant dans la barre d'outils image, positionnez le curseur dans la partie supérieure gauche de la fenêtre et cliquez, élargissez la fenêtre en glissant le curseur vers le coin inférieur droit et cliquez à nouveau. (Des côtés inférieurs à 1 mm. ne sont pas autorisés, ils ne contiendraient d'ailleurs pas le moindre caractère.)

Les fenêtres sont automatiquement triées par ordre de création: des flèches indiquent l'ordre de tri.

Vous pouvez également encadrer des blocs de texte "irréguliers" en dessinant des **fenêtres polygonales** autour d'eux. Les fenêtres non rectangulaires sont créées par l'assemblage de zones rectangulaires: dès que deux rectangles (du même



type) entrent en intersection, ils deviennent automatiquement une seule et même fenêtre! En quelque sorte, vous construisez une maison en ajoutant une pièce à l'autre...(Créer des fenêtres de tableaux polygonaux n'a pas de sens.)



Notez également que le fenêtrage manuel peut être combiné avec le tri des fenêtres: vous pouvez tracer de nouvelles fenêtres pendant que le “mode de tri” est activé. Vous utilisez alors le tri pour inclure certaines fenêtres détectées, et créer manuellement d’autres fenêtres là où l’analyse de page n’a pas obtenu les résultats idéaux. Dès que commencez à créer des fenêtres en “mode de tri”, toutes les fenêtres non sélectionnées sont aussitôt effacées!

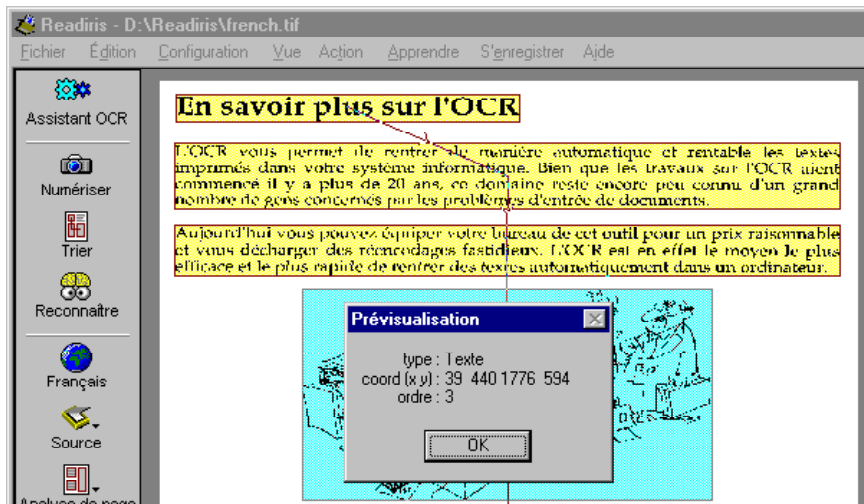
Pour modifier, déplacer ou supprimer des fenêtres, il vous faut d’abord les **sélectionner**. Pour ce faire, cliquez sur l’icône “Sélection Fenêtre” (la flèche) sur la barre d’outils image et cliquez à l’intérieur de la fenêtre. Des marqueurs rectangulaires apparaissent alors à chaque coin et sur les bords de la fenêtre.



## En savoir plus sur l'OCR

Pour **annuler la sélection** des fenêtres, cliquez ailleurs avec le bouton de la souris. Pour sélectionner des **fenêtres supplémentaires**, maintenez appuyée la touche Maj tout en cliquant sur ces fenêtres supplémentaires. Pour sélectionner une fenêtre et les **fenêtres incluses** (de l'autre type), maintenez enfoncée la touche Ctrl tout en cliquant sur la fenêtre principale.

Vous pouvez afficher le type, les coordonnées en pixel et l'ordre de tri d'une fenêtre en appuyant sur la touche Alt au moment de la sélection!



Voici pour la sélection des fenêtres. Pour **modifier** une fenêtre texte, sélectionnez-la, placez le curseur de la souris sur un marqueur et faites glisser le bord pour changer la taille de la fenêtre.



Pour **déplacer** une fenêtre, sélectionnez-la et déplacez-la simplement vers l'endroit désiré en gardant le bouton de la souris enfoncé.

Pour **supprimer** des fenêtres, sélectionnez-les et utilisez la commande "Couper" ou "Effacer" du menu "Édition". La commande "Couper" place votre fenêtre dans une mémoire tampon interne, la commande "Effacer" la supprime définitivement. Lorsque vous collez des fenêtres, elles sont réintégrées à leur place originale et vous devez les déplacer vers l'endroit souhaité.

En fait, *toutes* les commandes habituelles du menu "Édition" s'appliquent aux fenêtres: vous pouvez les supprimer, les couper, les copier et les coller! La commande "Annuler" peut également vous servir: si vous avez malencontreusement supprimé, déplacé, reformaté, etc. certaines fenêtres, "Annuler" annulera votre dernière opération!

Édition	
<u>Annuler</u>	Alt+BkSp
<u>C</u> ouper	Ctrl+X
<u>C</u> opier	Ctrl+C
<u>C</u> oller	Ctrl+V
<u>E</u> ffacer	Delete
Sélectionner <u>t</u> out	Ctrl+A

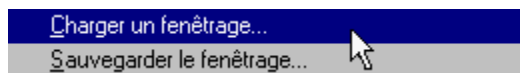
Notez aussi que des raccourcis existent pour toutes les commandes! Par exemple: pour supprimer toutes les fenêtres existantes, vous pouvez utiliser la commande "Sélectionner Tout" du menu "Édition" ou son raccourci clavier Ctrl+A et cliquer sur la commande "Effacer" ou son raccourci Supprimer. Vous pouvez maintenant recréer la mise en page désirée. Pour restaurer la mise en page antérieure, choisissez la commande "Annuler" ou le raccourci Alt+Backspace.



## **TROIS, ENREGISTRER DES MODÈLES DE FENÊTRAGE**

---

Le masque créé peut être enregistrée en **gabarits de fenêtrage** pour une utilisation ultérieure avec la commande "Sauvegarder le Fenêtrage" du menu "Fichier" et chargée en mémoire avec la commande "Charger un Fenêtrage".



Si vous devez reconnaître des documents avec une mise en page similaire, par exemple un rapport de 50 pages où les en-tête et pied de page doivent être supprimés pour des raisons évidentes, un simple modèle peut servir à aménager les 50 pages.

Lorsque vous chargez un fenêtrage fixe, l'analyse de page se désactive automatiquement et le gabarit choisi reste actif jusqu'à ce que vous réactiviez l'analyse de page sur la barre d'outils principale.

## **READIRIS VOUS EMMÈNE AUTOUR DU MONDE**

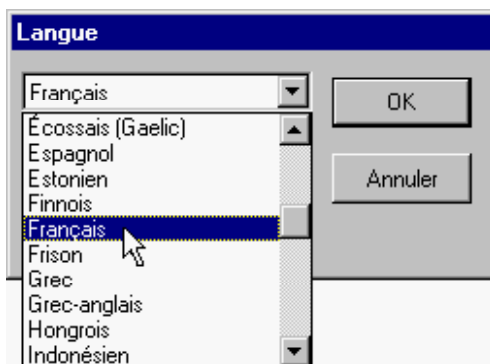
---

Maintenant que les fenêtres sont correctement définies, vous êtes presque prêt à effectuer la reconnaissance de caractères. "Presque" parce que nous n'avons pas encore vérifié la langue et les caractéristiques du document!

Le paramètre de langue se trouve sur la barre d'outils principale.



Cliquez sur le bouton "Langue" pour modifier la langue du document.



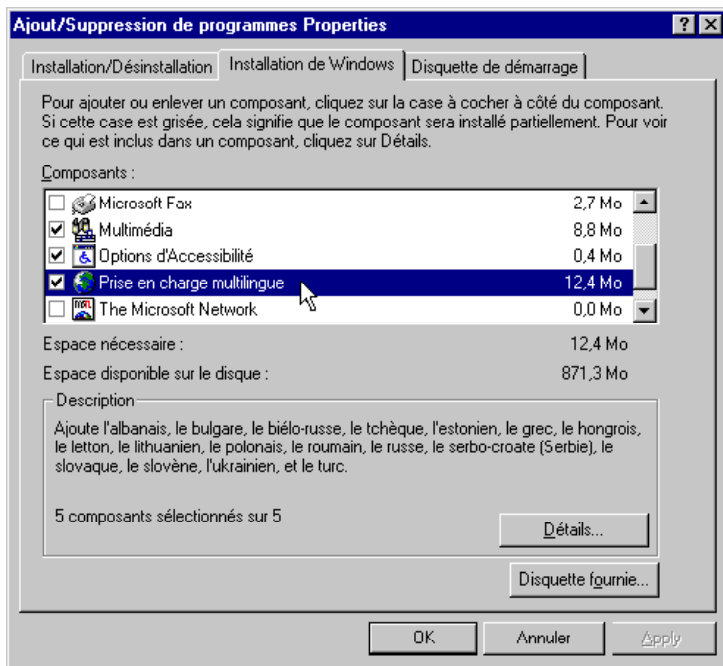
Vous pouvez appuyer sur une touche lettre pour la sélectionner directement: si le français est actuellement sélectionné et que vous souhaitez choisir le hongrois, vous pouvez enfoncer la lettre “H” de votre clavier: la langue hongroise sera sélectionnée. Quand différentes langues ont la même initiale, appuyez sur la lettre plusieurs fois afin de trouver les options. Par exemple: Readiris lit le polonais et le portugais. En appuyant sur “P” une fois, vous sélectionnez le polonais, en appuyant une seconde fois sur la touche “P”, vous sélectionnez le portugais, et en appuyant une troisième fois sur “P”, vous revenez au polonais. (Pour sélectionner une *autre* lettre, par exemple “D”, appuyez sur Arrière avant d’appuyer sur la touche “D”.)

Readiris est loin d’être limité au français: **55 langues** sont reconnues! Toutes les langues européennes et américaines sont lues, y compris les langues d’Europe centrale, le grec, le turc, les langues cyrilliques (“russes”) et les langues de l’Europe baltique.

De manière optionnelle, vous pouvez lire les documents asiatiques: l’option logicielle “Module d’OCR asiatique” offre la reconnaissance du japonais et du chinois simplifié. Le chinois traditionnel n’est pas supporté. (Le chinois simplifié est utilisé sur le continent chinois, le chinois traditionnel est utilisé par les “5 Grandes” communautés - Hong Kong, Taiwan, Singapore etc.)

Notez également que les variantes britanniques et américaines - “internationales” devrions-nous dire – de la langue anglaise sont distinguées.

Une configuration de Windows particulière est obligatoire pour afficher des caractères d'Europe centrale, grecs, turcs, cyrilliques et baltes. Vous pourriez avoir à installer le “composant” "Prise en Charge Multilingue" de Windows avant que votre système Windows ne puisse s'adapter à ces langues. Sélectionnez l'icône "Ajout/Suppression de Programmes" dans le "Panneau de Configuration" pour vérifier si ce module est installé sur votre PC. En fonction de la version du logiciel Readiris que vous avez acquise, le CD-ROM de Readiris peut contenir les informations nécessaires à l'installation de ce module de Windows.





Afin d’afficher et d’éditer des documents asiatiques, vous devez installer une version asiatique, japonaise ou chinoise du système Windows. Vous pouvez également utiliser des logiciels “d’émulation” spécialisés (tels que UnionWay AsianSuite ou TwinBridge AsianBridge) sur une version occidentale de Windows pour représenter correctement les idéogrammes de ces langues asiatiques.

Sélectionner correctement la langue du document est impératif. Basé sur la sélection d’une langue, le logiciel sait quel **jeu de symboles** il doit reconnaître. Un support multi-linguistique vous assure une reconnaissance correcte de caractères “exotiques” tels que ß, ñ, γ et ø.

Deuxièmement, le logiciel utilise de façon extensive ses **bases de données linguistiques** pour valider ses résultats. Supposez que vous ayez à lire le mot "président" où une tache d’encre donne au "r" l’aspect d’un "f". En consultant le lexique de la langue française, Readiris reconnaît de façon autonome le mot "président" et conclut qu’il n’y a aucun sens à la présence du symbole "f". Cette **technique “d’auto-apprentissage”** est évidemment déterminée par le contexte linguistique.

La linguistique offre une aide utile pour résoudre des **cas ambigus** tels qu’un "O" qui pourrait être confondu avec 'O'. Un autre cas typique est la lettre "l" et le chiffre '1' qui ont une forme identique dans de nombreuses fontes – pensez aux textes produits sur d’anciennes machines à écrire! Le contexte linguistique aide à déterminer s’il s’agit d’un "l" ou d’un '1'.

L’illustration ci-dessous montre différentes formes de "l" et de '1'. Les formes sur la première ligne ne sont pas ambiguës, les formes sur la deuxième ligne peuvent prêter à confusion, mais la linguistique peut résoudre leur interprétation. Lorsque le contexte ne suffit pas, l’utilisateur peut intervenir.

193 1950s. 1hr  
Well, Rossellini

## **READIRIS CHANGE DE LANGUE AU BESOIN**

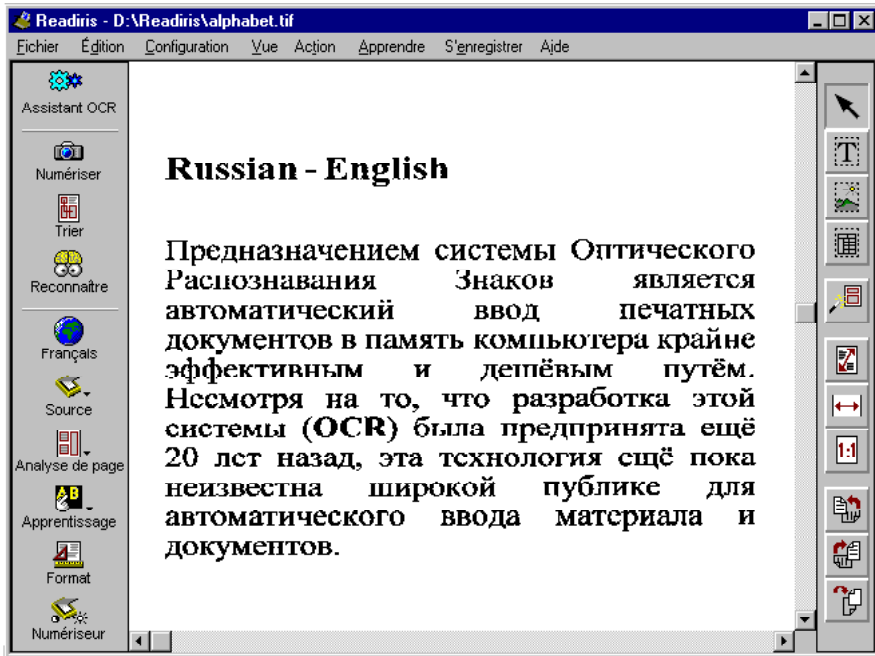
---

La performance ne s'arrête pas là: Readiris peut changer de langue en cours de phrase sans l'intervention de l'utilisateur! Quand des mots occidentaux apparaissent dans des documents grecs, cyrilliques ou asiatiques – beaucoup de noms propres non modifiables, des noms de marques, etc. sont écrits en utilisant les symboles latins -, Readiris peut s'adapter automatiquement au type d'alphabet. En d'autres mots, il est possible d'activer des **alphabets mélangés** de caractères grecs, cyrilliques ou asiatiques et occidentaux en même temps.

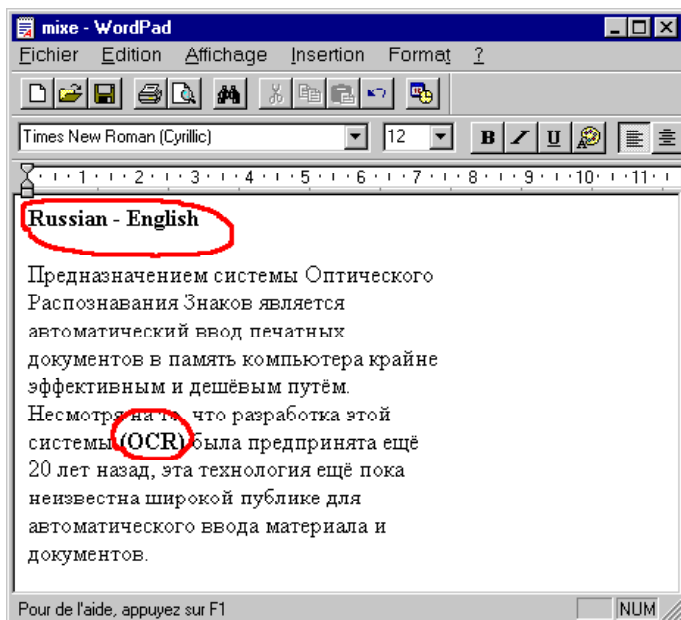
Assurez-vous de sélectionner "Grec-anglais" ou l'option correcte cyrillique – par exemple "Biélorusse-anglais". En d'autre mots: ne vous contentez pas juste de sélectionner "Grec" ou "Biélorusse" en espérant que les caractères occidentaux seront reconnus correctement!



Voici un exemple où un texte russe contient quelques mots anglais - ouvrez le fichier image ALPHABET.TIF si vous voulez essayer vous-même!



Le résultat final ressemble à ceci lorsque vous l'ouvrez avec votre traitement de texte – vous pourriez avoir à sélectionner la **fonte** cyrillique pour pouvoir afficher le texte russe correctement.



Pour **mélanger d'autres langues**, sélectionnez simplement la langue avec le jeu de caractères le plus étendu. Si vous avez un document où par exemple la traduction anglaise se trouve placée le long du texte français, vous devez sélectionner le français comme langue pour vous assurer que des caractères tels que ç, ù, et é seront correctement reconnus.

## DÉFINIR LES CARACTÉRISTIQUES DU DOCUMENT

Maintenant que la langue est définie, allons voir les autres caractéristiques du document. Vous pouvez peaufiner la reconnaissance en indiquant quelques spécifications du document: le type de fonte et l'espacement de caractère. (Ces



commandes ne s'appliquent pas aux documents asiatiques.) Qu'est-ce que cela veut dire?

Avant toute chose, indiquez à l'aide de la commande "Type de Police" dans le menu "Configuration" si vous voulez reconnaître des documents imprimés "normalement" ou en **impression matricielle**.

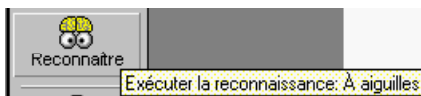


Il est nécessaire de choisir le paramètre correct: les symboles matriciels qualité "brouillon" sont isolés, composés de points séparés, et des routines de reconnaissance hautement spécialisées sont requises pour les reconnaître.

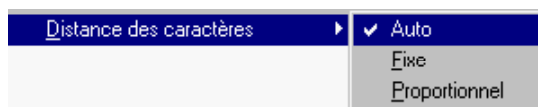
**ape-descended life**

La qualité d'impression matricielle qualité "lettre" exige l'option normale, de même que les **qualités d'impression** imprimées, tapées, impression laser et impression à jet d'encre. Evidemment, "Normal" est la valeur par défaut.

La bulle explicative du bouton "Reconnaître" indique le mode de lecture: lorsqu'aucun message n'est ajouté à la bulle explicative, le mode de lecture "normal" est utilisé, lorsque le message "A aiguilles" est ajouté à la bulle, le mode matriciel est d'application.



L'**espacement de caractère** peut être placé avec la commande "Distance des Caractères" sous le menu "Configuration".



Avec les fontes *fixes*, tous les symboles de la fonte ont la même largeur. Un "l" prend autant d'espace horizontal sur une ligne qu'un "w", comme c'est le cas dans cette phrase. Pensez



## GUIDE DE L'UTILISATEUR

aux documents produits en utilisant une machine à écrire, où le chariot déplace une distance fixe pour chaque symbole tapé.

Un espacement *proportionnel* signifie que la largeur d'un caractère dépend de sa forme. Les symboles comme "m" et "w" sont plus larges, occupent un espace horizontal plus large sur une ligne que les caractères minces "l" ou "j". Pratiquement tous les livres, magazines et journaux sont imprimés avec un espacement proportionnel.

Courier, une police fixe  
Letter Gothic. une police fixe

Century Schoolbook, une police proportionnelle  
Arial, une police proportionnelle

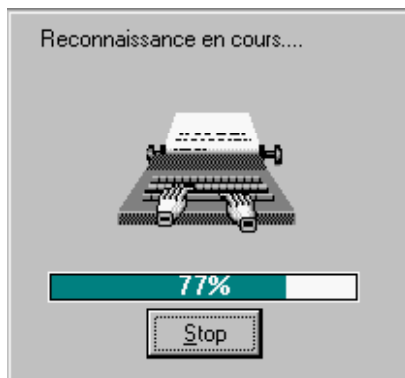
La solution la plus simple est de laisser cette option à tout moment sur la valeur par défaut "Auto": Readiris détectera la largeur de caractère automatiquement.

### **READIRIS DEVIENT CHAQUE FOIS PLUS INTELLIGENT!**

Quand la langue du document est choisie et ses caractéristiques définies, vous pouvez cliquer sur le bouton "Reconnaître".



La progression de l'OCR est indiquée sur l'écran. Vous pouvez cliquer le bouton "Stop" pour interrompre la reconnaissance du texte.



En fin de lecture, Readiris entre en phase interactive d'apprentissage si l'apprentissage est activé grâce au bouton "Apprentissage" sur la barre d'outils principale. L'apprentissage interactif est activé par défaut.

(L'apprentissage interactif ne s'applique pas aux documents asiatiques: l'apprentissage n'a pas de sens pour ces langues qui connaissent des milliers de symboles différents - en supposant que vous êtes capable d'introduire les idéogrammes, par exemple sur un clavier occidental!)



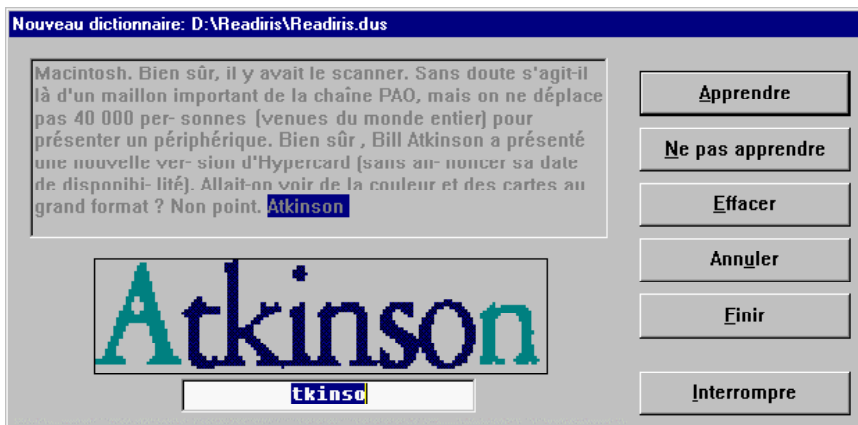
L'**apprentissage de fontes** peut sensiblement mettre en valeur l'exactitude du système d'identification. Quand l'utilisateur essaye de lire des symboles distordus et déformés comme on en trouve dans des documents réels ou des fontes stylisées que Readiris n'identifie pas de façon optimale, l'apprentissage peut surmonter ce "échec" provisoire.

L'apprentissage par l'utilisateur est également utilisé pour entraîner le système sur les **symboles spéciaux** que Readiris ne peut pas identifier, comme des symboles mathématiques et scientifiques et des dingbats. Quelques exemples: Readiris peut être formé pour identifier le symbole " $\pi$ " comme "pi" ou le caractère

## GUIDE DE L'UTILISATEUR

spécial "☎" en tant que "Tél.". (Cependant, la liste de symboles identifiés ne peut pas être étendue avec les symboles " $\pi$ " et "☎"!)

Le texte reconnu s'affiche progressivement et le système s'arrête sur les caractères douteux, ou - si vous traitez les caractères joints ("ligatures") - sur les chaînes de caractères douteuses. Elles sont toujours présentées dans leur contexte, les caractères douteux sont mis en surbrillance. Les caractères non reconnus sont représentés par un tilde (le symbole "~").



La première chose à faire est de vérifier si vous lancez le dictionnaire de fontes et le mode de dictionnaire corrects - ceux-ci sont toujours indiqués dans le titre de la fenêtre d'apprentissage. Si ce n'est pas le cas, cliquez sur le bouton "Interrompre" - l'image du document est à nouveau affichée avec son fenêtrage tel qu'il avait été créé - activez le bon dictionnaire ou mode de dictionnaire et relancez l'OCR. (Le fonctionnement des dictionnaires de fontes sera expliquée sous peu.)

Si nécessaire, entrez un caractère (ou une chaîne de caractères) pour une forme incorrecte ou inconnue et cliquez sur un des boutons suivants.



## Apprendre

Vous êtes d'accord avec la solution proposée ou vous la corrigez. Le programme sauvegarde le caractère douteux dans le dictionnaire de fontes comme étant "certifié", final. La reconnaissance suivante n'exigera plus votre intervention, la forme est considérée apprise une fois pour toutes.

Dans l'exemple ci-dessus, le système s'arrête sur des caractères collés, et nous cliquons sur "Apprendre" pour accepter une forme qui ne peut pas être confondue avec d'autres caractères.

## Ne Pas Apprendre

Vous êtes d'accord avec la solution proposée ou vous la corrigez. La différence avec le bouton "Apprendre" est que le symbole appris obtient le statut "incertain" dans le dictionnaire. Pour une future identification, le système proposera la solution apprise mais exigera toujours une confirmation.

Ce bouton est utilisé pour les symboles qui pourraient être confondus avec d'autres: un "e" déformé qui pourrait être confondu avec un "c", un "t" endommagé qui ressemblerait étroitement à un "r" etc.



## *GUIDE DE L'UTILISATEUR*

Le "e" ci-dessus est sérieusement endommagé - en fait il est proche du symbole "c" -, et vous devriez cliquer "Ne pas apprendre" pour ne pas le confondre avec le symbole "c".

### **Effacer**

On élimine la forme affichée du résultat. Ce bouton est utilisé pour ignorer les “bruits” sur les documents - taches de café etc. qui pourraient être reconnues comme points, virgule etc. -, et pour effacer tout autre symbole non désiré.

### **Annuler**

Vous retournez en arrière pour corriger les erreurs. Vous pouvez annuler les neuf dernières opération.

### **Finir**

L'apprentissage est interrompu mais l'OCR continue en mode automatique. Toutes les décisions du système sont acceptées sans validation de l'utilisateur.

Cliquez sur ce bouton quand vous voyez que la reconnaissance est précise et n'exige pas de correction approfondie.

Ne confondez pas le bouton "Finir" avec le bouton "Interrompre". Ce dernier arrête le processus alors qu'avec le bouton "Finir", le texte est créé, il ne subit simplement plus de correction approfondie!

## **LE RÔLE DES DICTIONNAIRES DE FONTES**

---

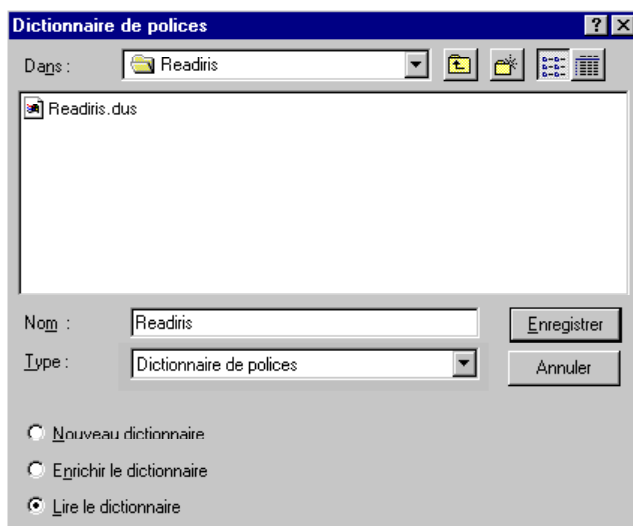
Les résultats de chaque session de formation sont temporairement maintenus dans la mémoire mais ils peuvent et devraient être enregistrés dans des fichiers appelés les “dictionnaires” pour une future utilisation.

Ces dictionnaires de polices doivent être chargés en mémoire quand vous voulez reconnaître des documents semblables afin d'utiliser l'intelligence supplémentaire qu'ils contiennent; de cette façon, Readiris tient compte de



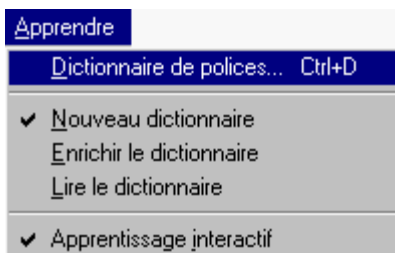
l'intelligence enregistrée dans ces bibliothèques de polices. Vous pourriez dire que Readiris devient toujours plus intelligent à chaque utilisation!

Comment cela se fait-il? L'enrichissement des dictionnaires de fontes est contrôlé par le menu "Apprendre": vous devez choisir un dictionnaire avec la commande "Dictionnaire de Polices" et déterminer son mode de fonctionnement.



Des **dictionnaires de polices** sont limités à 500 formes, et nous vous recommandons de créer des dictionnaires séparés pour des applications spécifiques, par exemple par type de document. Pour la clarté, donnez des noms significatifs aux dictionnaires de fonte, par exemple RAPPORT.DUS, PALATINO.DUS etc. Les dictionnaires ont par défaut l'extension \*.DUS. L'apprentissage et l'enrichissement n'ont plus d'effet quand le dictionnaire est rempli: les résultats de l'apprentissage ne sont plus mémorisés et stockés dans un dictionnaire.

Vous pouvez placer le mode dictionnaire sous la commande "Dictionnaire de Polices" ou directement sous le menu "Apprendre". Trois modes de dictionnaires sont disponibles: nouveau, enrichir et lire.



En choisissant "Nouveau Dictionnaire", vous indiquez que les résultats d'apprentissage seront sauvegardés dans un *nouveau* dictionnaire. (Si vous choisissez un dictionnaire existant, son contenu sera effacé.)

Le mode d'enrichissement indique que les résultats d'apprentissage seront sauvegardés dans un dictionnaire *existant*: la reconnaissance se sert de l'intelligence déjà contenue dans le dictionnaire, et vous lui ajoutez les nouvelles formes de fonte. En termes simples, cette option vous permet de construire un dictionnaire de fontes en plusieurs étapes.

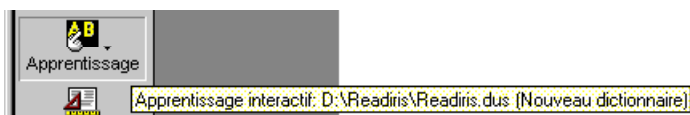
(Quand vous définissez le nom de fichier d'un nouveau dictionnaire et activez le mode d'ajout, un dictionnaire vide est créé et vous le complétez.)

Avec la dernière option, "Lire le Dictionnaire", le dictionnaire fonctionne en mode inaltérable: vous vous servez du dictionnaire *sans* lui ajouter une nouvelle forme de fonte.

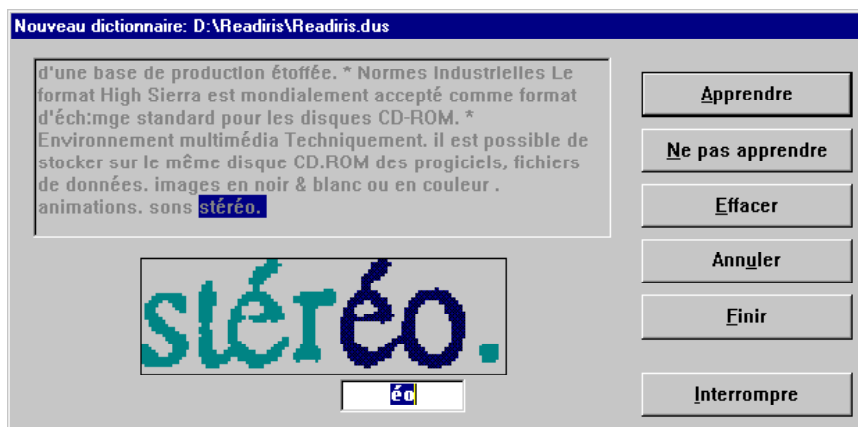
Choisissez le nouveau mode quand une page simple est reconnue. Pour lire beaucoup de pages du même type - des pages avec une qualité d'impression et des fontes similaires - choisissez le mode nouveau pour la première page, le mode ajouter pour les pages suivantes et le mode lecture pour le reste du document. Quand des documents de plusieurs pages sont reconnus, la numérisation d'une page supplémentaire met automatiquement le dictionnaire d'utilisateur en mode d'ajout.



Sachez que votre bouton d'apprentissage vous indiquera à tout moment grâce à sa bulle explicative quel est le dictionnaire de fontes actif et dans quel mode ce dictionnaire fonctionne.



Quand vous êtes en mode d'apprentissage interactif, le dictionnaire et son mode opératoire sont indiqués en titre de fenêtre; en cas d'erreur, vous pouvez cliquer sur "Interrompre" et recommencer.



## ENVOYER LE RÉSULTAT DIRECTEMENT DANS VOTRE APPLICATION

L'apprentissage interactif vient en conclusion à la reconnaissance de caractères. Comme Microsoft Word 97 fonctionne comme application cible par défaut, votre traitement de texte est démarré automatiquement (si nécessaire) à la fin de la reconnaissance, et le texte reconnu est inséré.



## GUIDE DE L'UTILISATEUR

L'image numérisée est à nouveau affichée avec son fenêtrage. L'image est disponible pour la poursuite du traitement, elle est affichée là jusqu'à ce que vous numérisiez une autre page.

Vous venez en effet de convertir un document papier en un fichier ordinateur éditable, et ce 10 à 25 fois plus vite qu'en dactylographie manuelle! Vous pouvez maintenant comparer votre texte à l'image qui se trouve dans votre fenêtre Readiris.

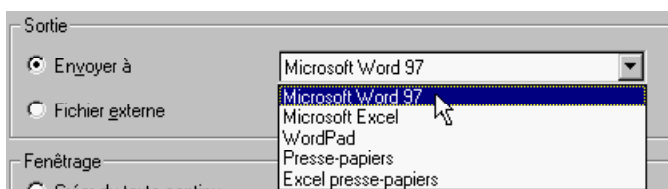
En fait, Readiris vous offre trois méthodes pour enregistrer le résultat OCR: envoyer le document reconnu dans une application cible, sauvegarder les fichiers dans un fichier externe et placer les résultats dans le presse-papiers Windows.

La cible sortie est sélectionnée avec le bouton "Format" (ou la commande "Format de Texte" du menu "Configuration").





L'option "Envoyer à" offre un lien OCR direct entre votre numériseur et vos applications Windows: vous **envoyez** les documents numérisés directement dans votre traitement de texte ou dans votre tableur! Readiris exporte ses données directement dans Microsoft Word 97, Microsoft Excel et l'éditeur Windows WordPad.



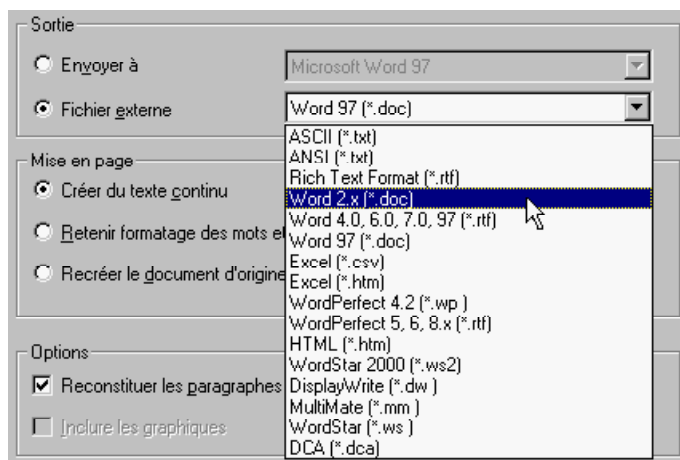
À la fin de la reconnaissance, l'application cible démarre automatiquement et le document reconnu est inséré dans un nouveau fichier texte ou feuille de calcul.

Veuillez patienter pendant le  
chargement de Microsoft Word 97

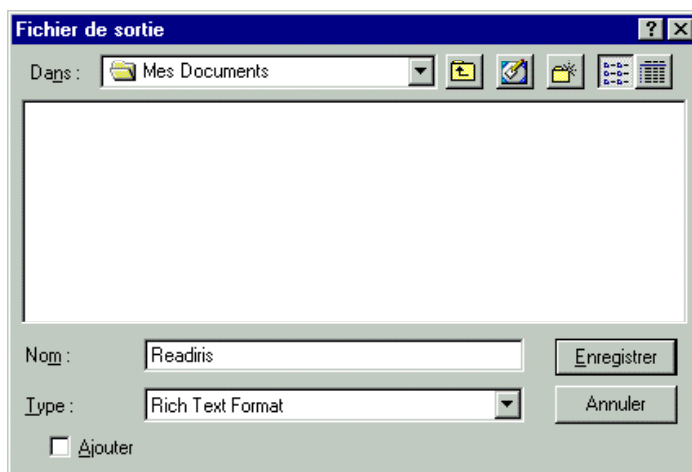
Cependant, n'oubliez pas que l'option "Envoyer à" permet également de copier le document reconnu dans le **presse-papiers** Windows; il n'y a donc pas de besoin impérieux d'exporter le résultat... ou de l'enregistrer dans un fichier externe!

## **SAUVEGARDER LES RÉSULTATS DANS UN FICHIER TEXTE**

Vous pouvez en effet placer le résultat du processus OCR dans un fichier "externe"! Readiris sort les données dans un grand nombre de traitements de texte et de tableurs - Microsoft Word (DOC), RTF and HTML, etc.



Quand vous sélectionnez l'option "Fichier externe", vous êtes invité à sauvegarder le texte à la fin de la reconnaissance.

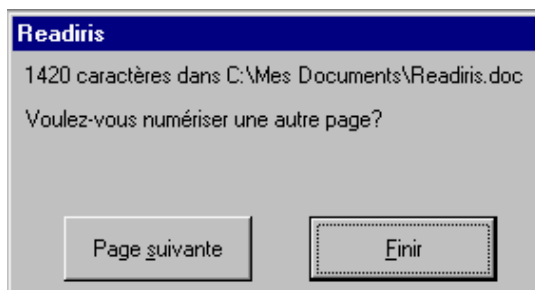


Vous pouvez créer un nouveau fichier et ajouter du texte à un fichier existant - c'est la raison pour laquelle il y a un bouton "Ajouter". Vous pourriez par exemple rajouter deux pages numérisées au texte d'une étude que vous auriez vous-même écrite. Si vous complétez un fichier existant, assurez-vous qu'il n'est pas ouvert, cela vous empêcherait d'écrire dans le fichier!



## RECONNAISSANCE DE PAGES MULTIPLES

Lorsque vous cliquez sur "Enregistrer" pour sauvegarder un texte, le système vous propose la lecture d'une autre page. (Si vous envoyez le texte dans une application cible ou dans le presse-papiers, vous ne serez évidemment pas invité à enregistrer le texte.)



A ce point, quelques brèves **statistiques** vous sont communiquées sur le nombre de caractères que vous venez de lire.

Vous pouvez maintenant ouvrir le texte reconnu avec votre traitement de texte ou votre éditeur de texte, l'importer dans le logiciel de publication (PAO) ou dans n'importe quelle application texte.

Cette leçon pour le traitement d'une page seule se termine ici. Mais comment sauvegarder le texte pour les pages supplémentaires? Comment traiter un document de plusieurs pages? C'est en fait très simple: cliquez sur "Page Suivante" quand vous êtes incité à lire une autre page. En le faisant, une autre page sera aussitôt numérisée si votre scanner fonctionne comme source d'image. Ensuite, le texte de sortie et le dictionnaire de fontes sont automatiquement mis en mode d'enrichissement pour vous permettre d'ajouter le texte et de poursuivre son traitement confortablement.

Vous pouvez poursuivre la reconnaissance d'autres pages: le nouveau texte est joint au fichier texte à chaque sauvegarde suivant la reconnaissance.



## ORGANISATION DU TEXTE SAISI

---

Sauvegarder ou exporter le texte signifie plus que sélectionner un mode de sortie ou donner un nom de fichier au texte reconnu. Vous sélectionnez également le format du fichier texte et déterminez l'apparence du texte reconnu. En bref, vous devez décider ce que vous allez faire du texte avant de lancer l'exécution.

Certaines options du bouton "Format" vous permettent de changer l'apparence du texte saisi.

Le **flux du texte** du document saisi est directement influencé par l'option "Reconstituer les Paragraphes" que vous trouverez sous le bouton "Format".



Laissez cette option activée afin que Readiris détecte les paragraphes: Readiris appliquera alors la **frappe en continu** comme dans les traitements de texte, sinon, un retour à la ligne est ajouté à la fin de chaque ligne et les mots avec un trait d'union restent tels quels. La détection des paragraphes est activée par défaut.

Voici un exemple: quand les trois premières lignes d'une colonne sont "Le nouveau prési-", "dent faisait signe du balcon." et "Sa femme l a rejoint.", la détection de paragraphe vous donne le résultat suivant "Le nouveau **président** faisait signe du balcon. **Sa** femme l a rejoint." Les deux parties du mot "président" ont été "recollée" et un espace ajouté à la fin de la première phrase, créant ainsi du texte courant.

Si la détection de paragraphe *n'a pas* été activée, la mise en page originale sera conservée, avec un retour à la ligne ajouté à la fin de chaque ligne.

(Le bouton "Format" a des options de formatage dont nous parlerons plus loin.)

## INSTALLATION DE VOTRE NUMÉRISEUR

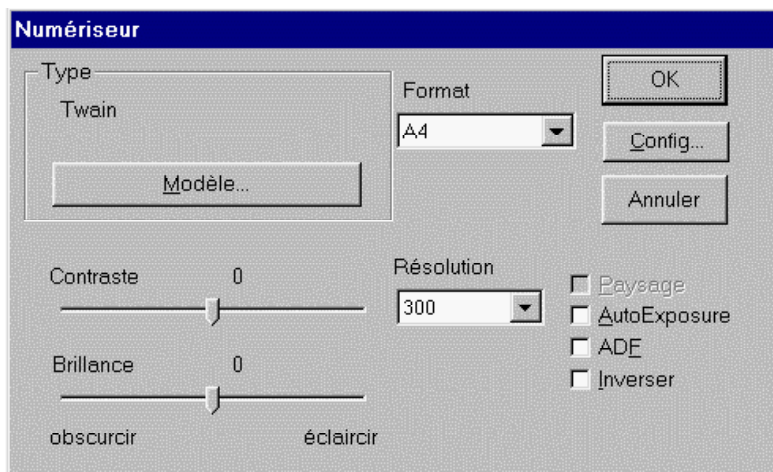
Déclarons maintenant le numériseur sous Readiris. Il est entendu que votre numériseur et ses pilotes sont correctement installés.

Si votre licence Readiris faisait partie des logiciels fournis avec le scanner, il se peut que cette étape ne soit pas nécessaire, dans la mesure où le numériseur peut déjà se trouver configuré sous Readiris.

Cliquez sur le bouton "Numériseur" de la barre d'outils principale.



Cliquez sur le bouton "Modèle" pour déterminer le **modèle** de votre numériseur.

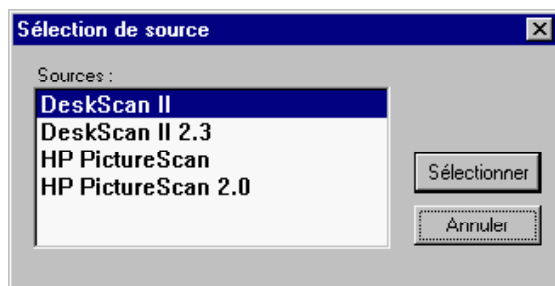


Quand vous sélectionnez l'option "<Image>" comme "numériseur", les images prés캔ées fonctionneront dans toutes les circonstances comme source d'image



– sans même devoir sélectionner la disquette comme source d'image en utilisant le bouton "Source" sur la barre d'outils principale.

Le bouton "Config." n'est opérationnel que si votre numériseur l'autorise. Il donne accès à certains paramètres pointus de numérisation; avec les scanners Twain, cliquer sur le bouton "Config." vous permet de choisir la source Twain. (Vous pouvez également utiliser la commande "Sélectionner Source" sous le menu "Fichier".)



Vous pouvez aussi avoir un scanner avec une **plate-forme de numérisation** comme Visioneer Paperport ou HP Document Assistant. En fonction de la version acquise du logiciel, vous pouvez également trouver sur le CD-ROM de Readiris des manuels spécialisés concernant les plates-formes de numérisation.

Une fois le scanner sélectionné, la même fenêtre peut vous permettre d'établir la résolution de numérisation, le format de page, l'orientation, la brillance et le contraste et peut vous permettre d'indiquer si vous allez utiliser l'alimentation du numériseur. Avec les numériseurs compatibles Twain, tous les paramètres de numérisation sont souvent définis dans l'interface Twain.

Sélectionnez une **résolution** de 300 ppp pour des applications normales, utilisez une résolution de 400 ppp pour des petites impressions (inférieures à 10 points) et quand le document est fort abîmé.

Readiris lit les caractères dont la **taille** va de 6 à 72 points (0,21 à 2,54 cm).



6 points

# 72 points

Readiris reconnaît aussi les “lettrines”, les caractères larges qui couvrent plusieurs lignes - ceux-ci ne pouvant évidemment pas dépasser 72 points!

**R**eadiris lit les  
“lettrines” (qui  
recouvrent plusieurs  
lignes) et les attribue  
à la première ligne du  
paragraphe.

En activant l’option "Paysage", vous indiquerez que le texte est orienté en largeur plutôt qu’en hauteur (documents “portraits”). L’orientation de page s’applique plutôt à des pages de format réduit: sur un scanner à plat A4, vous pouvez numériser des pages A5 (à moitié moins grandes) en mode paysage ou portrait, mais vous ne pouvez numériser toute la surface d’un A4 que dans un sens!



Lorsque vous voulez ajuster l’orientation de page, utilisez les outils de rotation sur la barre d’image. (Les commandes correspondantes sont dans le menu "Vue".)



Trois directions de rotation sont proposées: à gauche, à droite et 180°. La rotation prend quelques secondes: l'image elle-même est activée et non simplement son affichage à l'écran.

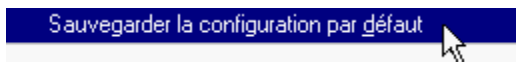
Définissez la **brillance** et, si disponible, le **contraste**. Le paramètre du contraste n'est disponible que sur certains numériseurs. Glissez la barre coulissante pour foncer ou éclaircir le document; les valeurs sélectionnées sont également représentées numériquement.

Enfin l'option "Inverser" vous permet de générer des **images "inversées"** - activez cette option pour les pages dont tout le texte est écrit en blanc sur fond noir.

## SAUVEGARDER LES PARAMÈTRES PAR DÉFAUT

---

Définissez correctement tous les paramètres par défaut et cliquez sur la commande "Sauvegarder la Configuration par Défaut" dans le menu "Fichier" pour une prochaine utilisation.

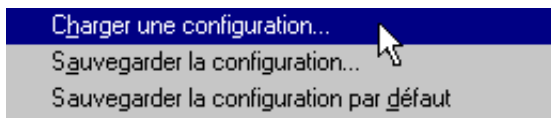


Les fichiers de paramètres contiennent plus que la **configuration** du numériseur: ils déterminent également si vous allez faire de l'apprentissage interactif, quel dictionnaire de fontes à utiliser (et son mode), quels sont la langue et le type de polices - par exemple une fonte normale, proportionnelle - des documents, quel mode et sortie et quelles options de formatage sont utilisés - par exemple du texte continu envoyé dans WordPad, etc. En bref, tous les paramètres opérationnels de Readiris sont stockés dans les fichiers de paramètres.

## SAUVEGARDE DE PARAMÈTRES SPÉCIFIQUES

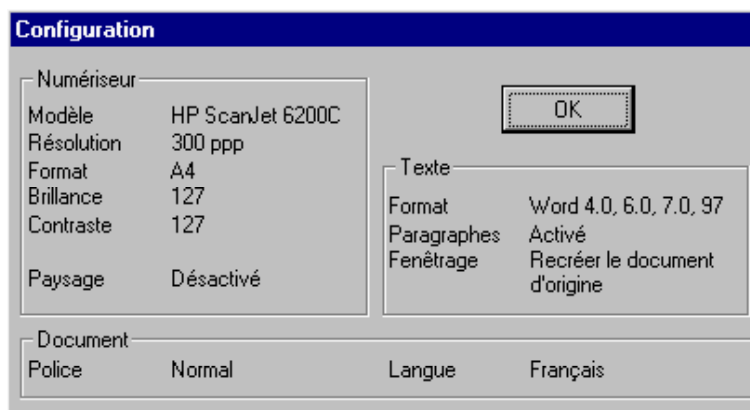
---

Les paramètres par défaut seront activés à chaque démarrage, mais vous pouvez sauvegarder des paramètres spécifiques afin de n'avoir pas à redéfinir les paramètres opérationnels. Les commandes "Sauvegarder la Configuration" et "Charger une Configuration" sous le menu "Fichier" s'en chargent.



Voici un exemple: si vous devez régulièrement faire de l'OCR sur des documents en français avec une mise en page spécifique, nous vous recommandons de créer un fichier de paramètres spécifique pour ce type de documents. Vous choisirez alors français comme langue du document, chargerez un gabarit de fenêtres pour n'avoir pas à réappliquer le même fenêtrage chaque fois, désactiver l'apprentissage mais activer un dictionnaire de fontes en mode lecture étant donné que des polices identiques sont utilisées à chaque fois, etc.

Si vous n'êtes pas certains des paramètres en cours, vous ne devez pas "plonger" à l'intérieur de tous les menus et commandes. Utilisez la commande "Informations" du menu "Fichier" pour savoir ce qu'il en est.



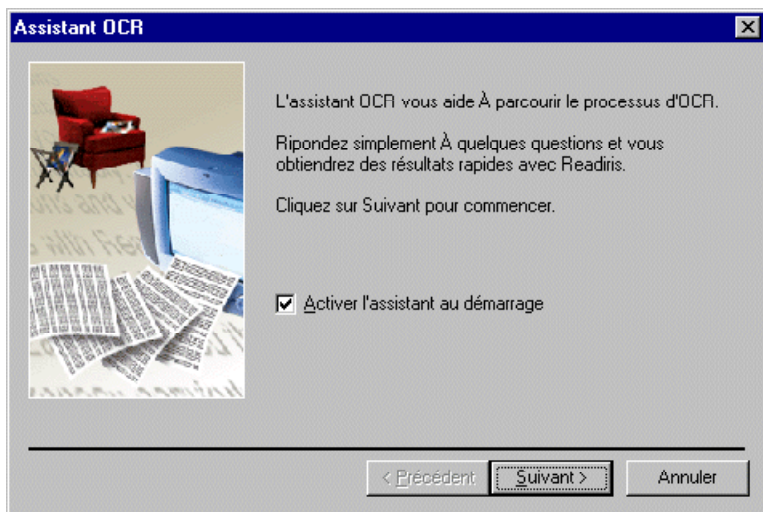
## FAIRE TRAVAILLER L'ASSISTANT

Maintenant que le numériseur est installé, nous pouvons entamer la saisie des documents. Plutôt que de parcourir tous les paramètres et options, nous allons nous servir de l'assistant OCR, une manière très confortable de reconnaître des pages.

Cliquez sur le bouton "Assistant OCR" dans la barre d'outils principale (ou sélectionnez la commande "Assistant OCR" dans le menu "Action").



L'assistant vous mène à travers le processus d'OCR de manière confortable: vous répondez à quelques questions simples pour obtenir des résultats rapides avec Readiris.



Notez que cet assistant démarre à chaque fois que vous lancez Readiris; afin d'éviter ce phénomène, désactivez l'option "Activer l'Assistant au Démarrage" dans le premier écran de l'assistant (ou utilisez l'option équivalente sous le menu "Configuration").

## RECONNAISSANCE AUTOMATIQUE DES PAGES

L'assistant OCR constitue une manière sémi-automatique d'exécuter l'OCR. Readiris offre également la reconnaissance *entièrement* automatique: plutôt que de parcourir les étapes pas à pas, nous allons reconnaître une page immédiatement, sans aucune interruption.

Pour remplacer le bouton "Assistant OCR" par le bouton "Auto", désactivez l'option "Placer l'Assistant sur la Barre des Boutons" sous le menu "Configuration".



✓ Placer l'assistant sur la barre des boutons



Sélectionnez le numériseur comme source d'images à l'aide du bouton "Source" et cliquez sur le bouton "Auto" (ou sélectionnez la commande "OCR Automatique" sous le menu "Action").



La **numérisation automatique** signifie qu'une page est successivement numérisée, fenêtrée par analyse de page ou sur base d'un gabarit de fenêtres et reconnue sans apprentissage interactif. La seule chose que vous ayez à faire est de lancer la numérisation et enregistrer le texte reconnu, les étapes intermédiaires sont assurées par Readiris.

## **READIRIS RECRÉE LA MISE EN PAGE D'UN DOCUMENT**

La reconnaissance automatique, qui rend le processus de reconnaissance autonome, ne doit pas être confondue avec de "l'autoformatage"! Autoformatage signifie que Readiris recrée une **copie facsimile** du document numérisé: le formatage des mots, des paragraphes et de la page de votre document original sont appliqués.

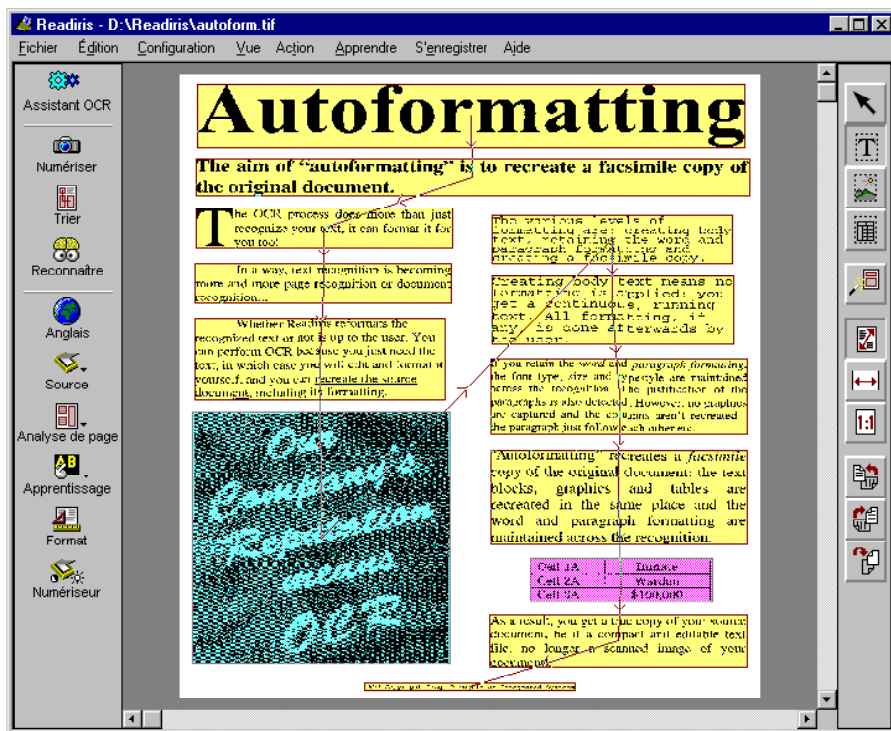
Des polices similaires (serif et sans serif, proportionnel et fixes, normal et condensé) sont utilisés comme dans le document source; la taille des points, le style de caractères (gras, italique et souligné) sont maintenus au cours de la reconnaissance. La tabulation et l'alignement (à gauche, centré, à droite et justifié) de chaque bloc de texte sont recréés. L'emplacement des colonnes, des parties de texte et des graphiques imitent le document original.

Readiris vous permet donc d'archiver une véritable copie de votre document original, mais il s'agit cette fois d'un texte éditable ou d'un fichier texte compact et non pas d'une image numérisée!

## GUIDE DE L'UTILISATEUR

Tout ceci implique qu'un tri des fenêtres exécuté par l'utilisateur ne s'applique que *partiellement* quand "l'autoformatage" est utilisé: vous pouvez inclure et exclure certaines zones, mais quel que soit le tri fait sur les résultats de l'analyse de page, il est ignoré!

Voici un exemple du mode de fonctionnement. Pour vous familiariser avec cette caractéristique, ouvrez l'image AUTOFORM.TIF dans votre dossier Readiris.





Cliquez sur le bouton "Format" dans la barre d'outils principale et envoyez le résultat de la lecture dans Microsoft Word ou sélectionnez le format "RTF" (Rich Text Format) ou Word (DOC) comme format à utiliser. Deuxièmement, activez "Recréer le Document d'Origine" comme option de mise en page. (L'option "Reconstituer les Paragraphes" est activé par défaut afin d'appliquer la frappe continue à l'intérieur des paragraphes.)



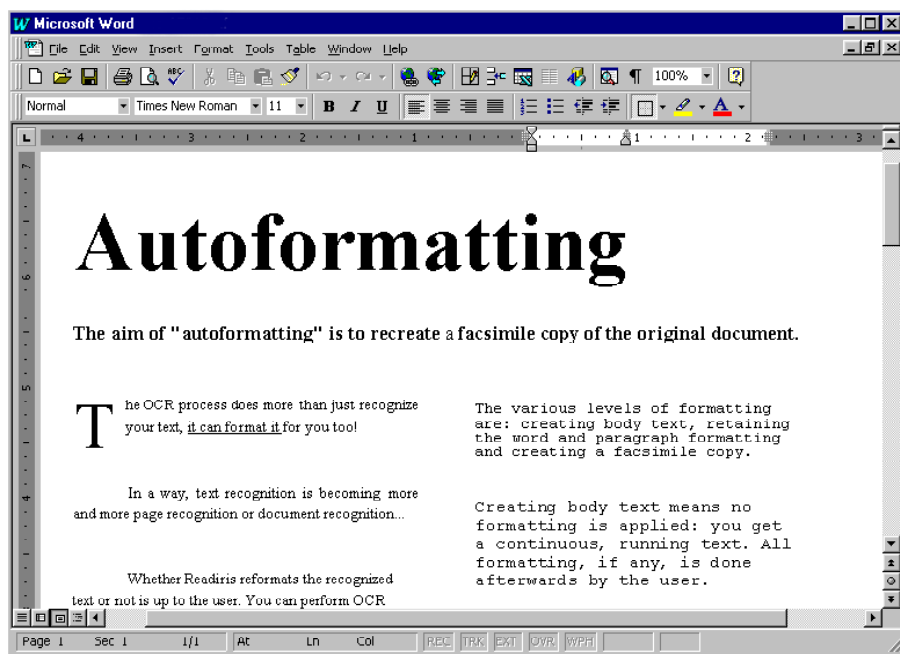
Remarquez que la reconstruction de la mise en page est limitée aux formats Word (DOC) et RTF et aux documents envoyés dans Microsoft Word. Les formats Word (DOC) et RTF sont largement réappandus et peuvent être ouverts par n'importe quel traitement de texte populaire.

Des autres formats, "pauvres", générant du texte simple tels que ASCII, ANSI, etc. ne supportent pas les codes de formatages avancés et ne peuvent donc pas fournir "l'autoformatage". De même, sachez que WordPad est un éditeur de texte

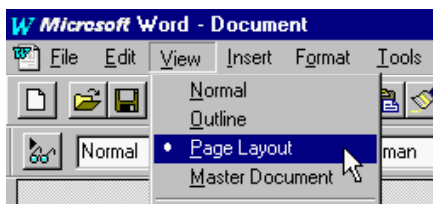


“réduit” et non un traitement de texte avancé; WordPad peut ouvrir des fichiers RTF mais ignore la plupart des éléments de formatage tels que les blocs de texte, l'alignement, etc.

Lorsque le texte reconnu est ouvert par un traitement de texte, il ressemble à ceci sans la *moindre* intervention de l'utilisateur.



Pour bien voir le résultat, vous devez activer le mode “WYSIWIG” de votre traitement de texte, plus souvent appelé “mise en page”. Cependant, si vous envoyez le document reconnu directement dans Microsoft Word 97, la vue “mise en page” sera activée automatiquement!

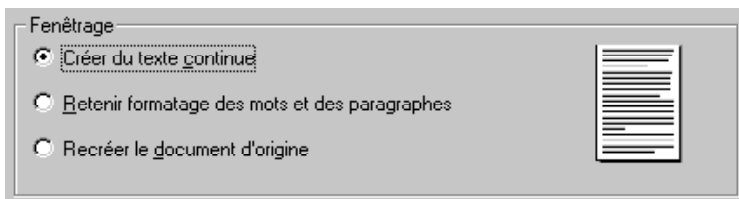


En résumé, Readiris ne reconnaît pas seulement vos textes, il est aussi capable de les formater pour vous. L'OCR n'est plus seulement de la reconnaissance de texte, cela devient plus de la **reconnaissance** de **page** ou de document!

## FORMATER LE TEXTE, DEUXIÈME PARTIE

Les autres options de mise en page sont "Créer du Texte Continu" et "Retenir Formatage des Mots et des Paragraphes".

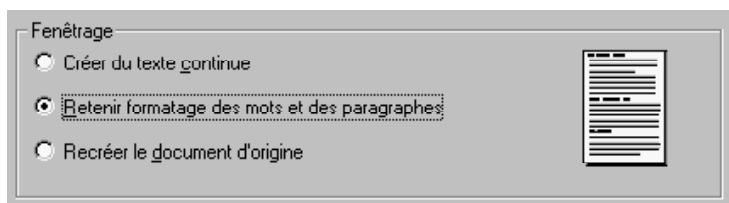
Comme l'illustre l'icône à la droite de la fenêtre, créer un corps de texte signifie que vous créez un **texte "courant"** non formaté. Le texte est saisi mais sa mise en forme est totalement ignorée. Utilisez cette option lorsque vous avez besoin de resaisir un texte mais pas sa mise en page.



(Générer des textes asiatiques implique la création d'un texte continu - il est impossible d'appliquer "l'autoformatage" ou de retenir le formatage des mots et des paragraphes. Le nombre de formats de fichier est également réduit.)

L'option "Retenir Formatage des Mots et des Paragraphes" permet une "demi-mesure": le **formatage de mots** - type de fonte (serif et sans serif, proportionnel et fixe, normal et condensé), la taille des points et le style de caractères (gras,

italique et souligné) sont maintenus au cours de la reconnaissance. Il en est de même pour le **formatage des paragraphes** - la tabulation et l'alignement (à gauche, centré, à droite et justifié).



Ne confondez pas cette option de formatage avec l'option "d'autoformatage" complète: cette option place simplement les paragraphes les uns à la suite des autres, elle ne créera pas de colonnes ou ne copiera pas l'emplacement relatif des blocs. Elle ne réintègrera pas non plus les graphiques, comme "l'autoformatage" peut le faire.

## **SAUVEGARDER DES GRAPHIQUES**

---

Dans notre exemple, le graphique était inclus dans le texte reconnu: cela dépend de l'option de formatage "Inclure les Graphiques". Sauvegarder des graphiques à l'intérieur d'un texte n'est possible que grâce à l'option d'autoformatage "complète", et donc pas avec des formats de texte "pauvres" tels qu'ASCII, ANSI, etc.



Mais avec Readiris, vous pouvez aussi sauvegarder des graphiques sans effectuer de la reconnaissance de texte. Comme le logiciel OCR de Readiris génère des **images en noir et blanc**, mais pas des images en niveaux de gris ou



en couleur, numériser des photos avec Readiris n'a pas de sens, par contre vous pouvez numériser des graphiques en noir et blanc sans problème.

Pour ce faire, dessinez une fenêtre graphique autour de l'illustration. Ceci se fait de la même manière que pour dessiner des fenêtres de texte et tableau, simplement en sélectionnant l'outil "Fenêtre Graphique" cette fois-ci.



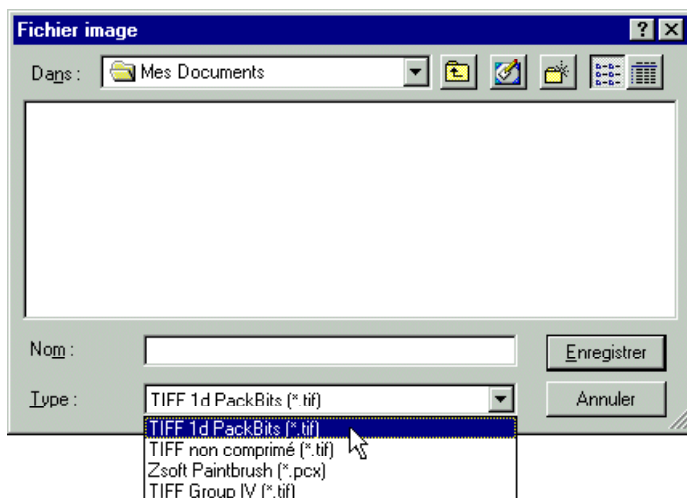
(Comme les autres types de fenêtres, la bulle explicative de l'outil "Fenêtre Graphique" vous dira le nombre de fenêtre graphiques sur l'écran. Et vous pouvez enfoncer la touche Alt lorsque vous sélectionnez une fenêtre afin d'en afficher le type et les coordonnées.)

Sélectionnez maintenant la commande "Sauvegarder les Graphiques" sous le menu "Fichier".



Sauvegarder la pleine page comme image...

Vous êtes invité à donner un nom de fichier.



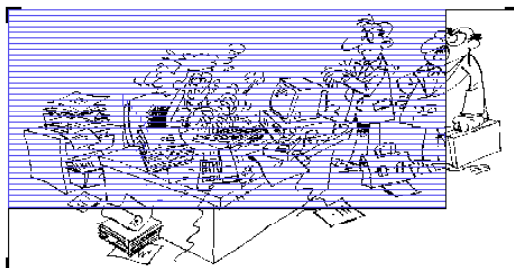
Déterminez quel format de fichier graphique vous souhaitez utiliser. Sélectionnez un format qui est supporté par votre logiciel de retouche photo. Les formats TIFF et Paintbrush (PCX) sont disponibles. (Des fichiers Paintbrush peuvent toujours être ouverts à partir de l'accessoire de Windows Paint.)



Les graphiques sont sauvegardés dans un seul fichier. Vous ne devez pas vous limiter à un seul fichier, mais si vous dessinez plusieurs fenêtres graphiques, elles seront rassemblées dans un seul fichier. (Vous pouvez utiliser la commande de découpe de votre logiciel de retouche photo pour les séparer.)

Des côtés inférieurs à 1 mm ne sont pas autorisés – des bitmaps de cette taille peuvent difficilement contenir la moindre information. Les fenêtres “irrégulières”, non rectangulaires sont transformées en rectangles invisibles: Readiris couvre la plus grande surface établie sur les coordonnées maximales. Dans l’exemple ci-dessous, la fenêtre graphique entourée d’un bord noir est complétée de la zone hachurée.

ur computer automatically.



ent is read by your scanner. This device acts as the "eye" of yo

## LECTURE DE FAX ET RECONNAISSANCE EN DIFFÉRÉ

Sauvegarder des images comme fichier image offre une autre possibilité: vous pouvez sauvegarder la page *entière* et y effectuer l'**OCR en différé**. C'est ce que nous avons fait avec les images préscannées au cours des didacticiels.

Numérisez le document et choisissez la commande "Sauvegarder la Page Pleine Comme Image" du menu "Fichier". Vous serez invité à enregistrer l'image entière dans un fichier TIFF ou Paintbrush (PCX).

Sauvegarder la pleine page comme image...



Vous pouvez maintenant sélectionner la disquette comme source d'images et ouvrir le fichier image avec le bouton "Ouvrir" (ou à l'aide de la commande similaire sous le menu "Action"). (Si vous utilisez la commande "Ouvrir" dans le menu "Fichier", vous ne devrez pas actualiser la source d'image.)

Readiris donne la possibilité d'ouvrir des images TIFF en noir et blanc (non-comprimées, comprimées en packbits, Groupe 3 et Groupe 4), des images Paintbrush (PCX) et des formats Windows bitmaps (BMP). Cette fonctionnalité est particulièrement utile pour convertir vos **télécopies** en textes éditables!



Si vous avez une quelconque influence sur vos correspondants, demandez-leur de vous envoyer des télécopies en utilisant la résolution “fine” – ces fax ont une résolution supérieure à 200 ppp et obtiendront de meilleurs résultats d’OCR.

N’oubliez pas que vous pouvez “**glisser-déposer**” des images pour les ouvrir: faites glisser une image de l’Explorateur de Windows vers la zone image de Readiris et elle est immédiatement ouverte!

## **RECONNAISSANCE DE TABLEAUX**

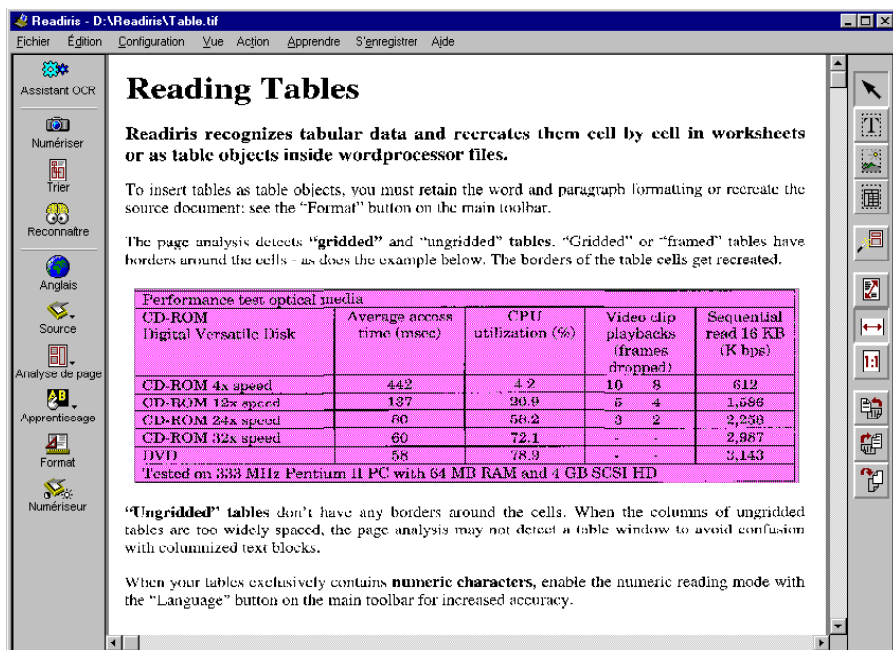
---

Nous avons jusqu’ici reconnu des textes, des fax et sauvegardé des graphiques. Traitons maintenant un tableau. Prenez un tableau et numérisez-le ou ouvrez l’exemple en image de votre dossier Readiris, TABLE.TIF.

En fait, l’image TABLE.TIF contient deux tableaux et ce n’est pas un hasard! L’analyse de page détecte les deux tableaux comme zone de tableau, et Readiris les reconstruira en recréant un tableau cellule par cellule dans votre feuille de calcul ou en insérant un objet tableau dans votre traitement de texte.

Découvrons ensemble les nombreuses possibilités, en commençant par le tableau encadré. Un tableau “encadré” a des bordures autour de chaque cellule.





Effectuez la reconnaissance avec l'option de mise en page "Retenir Formatage des Mots et des Paragraphes" ou l'option "Recréer le Document d'Origine" et le tableau est recréé. Ouvrez votre traitement (si vous n'avez pas envoyé le résultat directement dans Word) afin de vérifier la sortie. (Vous pourriez aussi bien avoir inclus le texte en cours dans le fichier texte.)



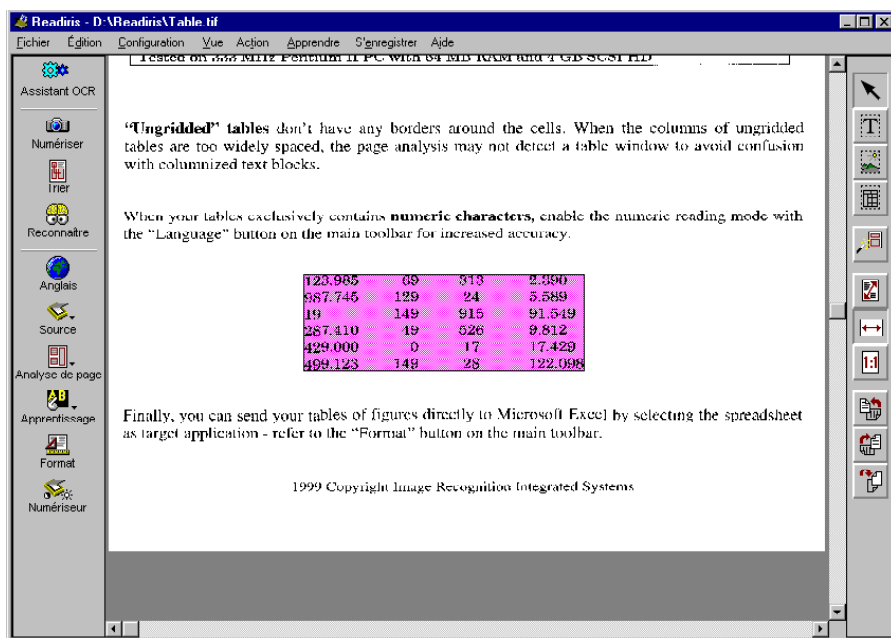
The screenshot shows the Microsoft Word 97 interface. The title bar reads 'Microsoft Word - table'. The menu bar includes File, Edit, View, Insert, Format, Tools, Table, Window, and Help. The toolbar contains various icons for file operations, editing, and formatting. The status bar at the bottom shows 'Page 1', 'Sec 1', '1/1', 'At 8.5cm', 'Ln 13', 'Col 1', and a row of icons for spell check, mark, extend, overtype, and word wrap.

Tip of the Day: To hyphenate a document as you type, click Hyphenation on the Tools menu.

Performance test optical media					
CD-ROM Digital Versatile Disk	Average access time (msec)	CPU utilization (%)	Video clip playbacks (frames dropped)		Sequential read 16 KB (K bps)
CD-ROM 4x speed	442	4.2	10	8	612
CD-ROM 12x speed	137	20.9	5	4	1,586
CD-ROM 24x speed	80	58.2	3	2	2,258
CD-ROM 32x speed	60	72.1	-	-	2,987
DVD	58	78.9	-	-	3,143

Tested on 333 MHz Pentium II PC with 64 MB RAM and 4 GB SCSI HD

Maintenant l'exemple où le tableau n'a pas de bordure autour des cellules. Remarquez que l'analyse de page détecte néanmoins le tableau!



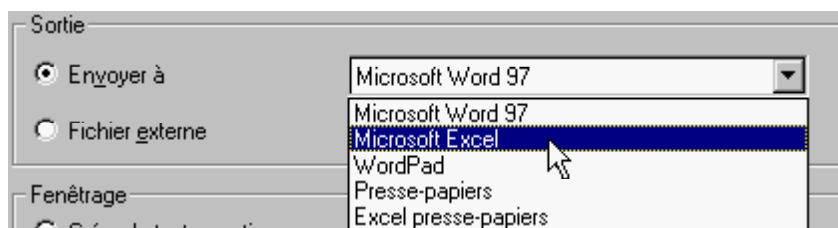
Pour obtenir une exactitude optimale d'OCR, vous devriez limiter la reconnaissance aux **symboles numériques** avec le bouton "Langue". (Le mode numérique n'est pas strictement numérique, il inclut les symboles 0 à 9, +, \*, :, %, , (virgule), . (point), (, ), -, =, \$ et £.)



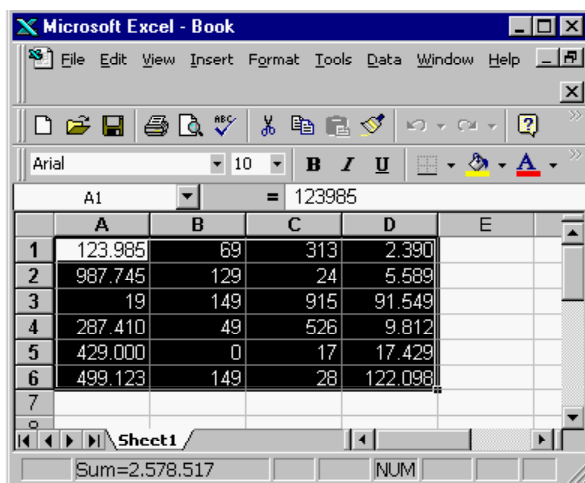
Comme vous ne pouvez faire ceci que lorsque le tableau ne contient pas de symboles alphabétiques – autrement les portions de texte ne seront pas reconnues correctement – nous pouvons maintenant activer le mode numérique mais ne pourrions pas le faire pour le premier tableau.



Cette fois-ci, nous enverrons le résultat directement dans le tableur Microsoft Excel; nous sélectionnons donc Excel comme application cible à l'aide du bouton "Format".



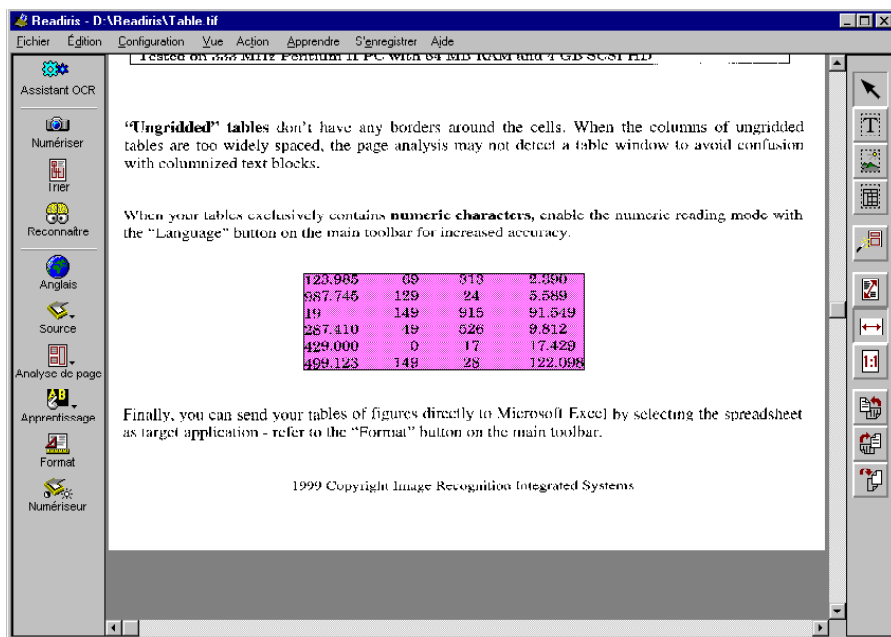
Le tableur est lancé automatiquement et la sortie se présente de cette façon: la structure typique de tableau avec lignes et colonnes est recrée, et vous pouvez commencer immédiatement le traitement des données.



Ouvrez votre traitement de texte pour avoir un aperçu du résultat – vous pouvez utiliser la commande "Convertir Texte en Tableau" de votre traitement de texte

pour finaliser les choses. (Vous pourriez aussi bien avoir inclus le texte en cours dans le fichier texte.)

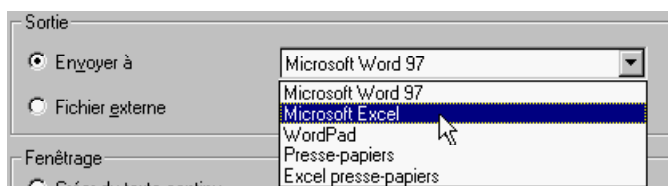
Maintenant l'exemple où le tableau n'a *pas de bordure autour des cellules*. Comme l'analyse de page ne détecte pas le tableau comme tel, l'utilisateur doit dessiner des fenêtres de texte autour de chaque colonne du tableau afin d'obtenir une apparence nette correspondant aux cellules.



Cette fois, vous devez sélectionner un format de tableau à l'aide du bouton "Format" pour ajouter des codes de tabulation entre les cellules. (Les formats de texte "généraux" *ne* suffisent *pas*, il faut un format spécifique pour créer les lignes et les colonnes!)



Comme avec les documents textes, vous pouvez créer un fichier externe qui peut être importé dans votre feuille de calcul, ou copier directement le tableau dans le presse-papiers et le copier directement dans votre feuille de calcul.

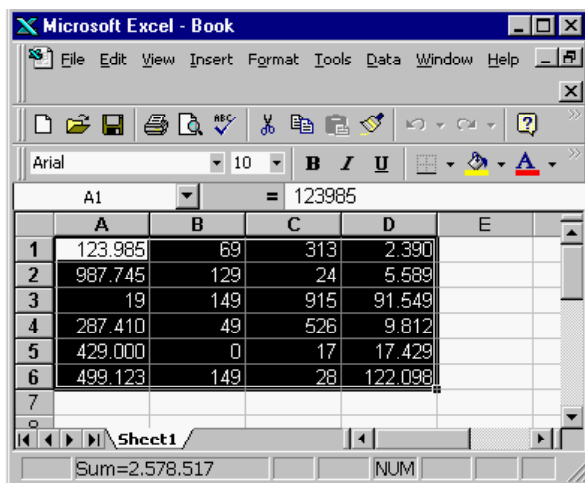


Pour obtenir une exactitude optimale d'OCR, vous devriez limiter la reconnaissance aux **symboles numériques** avec le bouton "Langue". (Le mode numérique n'est pas strictement numérique, il inclut les symboles 0 à 9, +, \*, :, %, , (virgule), . (point), (, ), -, =, \$ et £.)



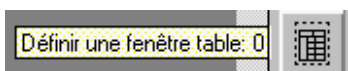
Comme vous ne pouvez faire ceci que lorsque le tableau ne contient pas de symboles alphabétiques – autrement les portions de texte ne seront pas reconnues correctement - nous pouvons maintenant activer le mode numérique mais ne pourrions pas le faire pour le premier tableau.

Le résultat, lorsqu'il est ouvert ou copié dans votre feuille de calcul, ressemble à ceci: une structure typique de tableau avec lignes et colonnes est recrée et vous pouvez commencer immédiatement le traitement des données.



Notez également que vous pouvez copier les tableaux dans le presse-papiers - ce qui explique la présence de l'option "Envoyer à - Presse-papiers Excel" et vous pouvez les enregistrer dans un fichier externe - d'où le format de fichier Excel ("CSV" signifie les "données délimitées par une virgule").

Il se peut que l'analyse de page ne détecte pas certains tableaux non encadrés comme des fenêtres de tableaux; ceci peut arriver dans le cas où les colonnes sont trop espacées. Readiris essaye en effet d'éviter la confusion avec des blocs de texte organisés en colonnes. Afin de créer manuellement une fenêtre de tableau, cliquez sur l'outil "Fenêtre de Tableau" dans la barre des outils image et créez la fenêtre normalement. La bulle explicative du bouton indiquera le nombre de fenêtres tableau.





## LA FONCTION “CONNECT” OFFRE UN LIEN DIRECT

---

En fait, il y a un autre moyen de reconnaître les tableaux: vous pouvez le faire à partir de votre feuille de calcul!

La fonctionnalité “Connect” offre un lien OCR direct entre votre numériseur et vos applications Windows – que ce soit le traitement de texte ou la feuille de calcul!

Comme il y a une série d’étapes manuelles qu’aucune procédure d’**installation** ne peut prendre en charge, vous devez faire appel à l’aide en ligne ou au fichier CONREAD.DOC dans votre dossier Readiris pour apprendre comment l’installer – il faut une macro pour utiliser la fonction “Connect” à partir de Word, Excel et WordPerfect. (Ce fichier du type “Lisez-moi” est affiché automatiquement en fin d’installation du logiciel.)

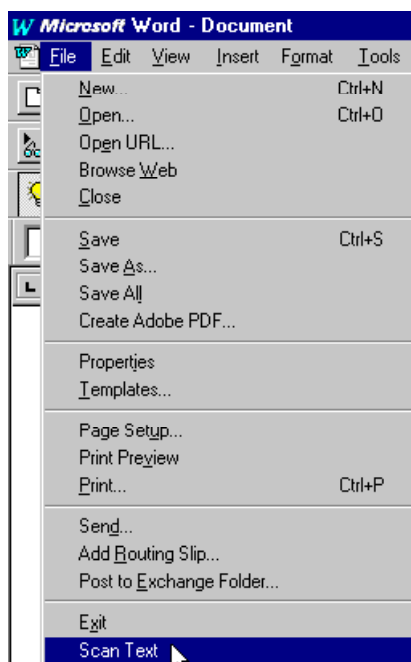




Une fois qu'elle est correctement installée, la fonction "Connect" ajoute une nouvelle commande aux menus de votre application. Dans les exemples ci-



dessous, la commande "Numériser Texte" est ajoutée au menu "Fichier" de Word, et le menu "Numériser Tableau" est ajouté à la barre des menus d'Excel.



Voici comment cela fonctionne: Readiris *ne peut pas* être lancé lorsque vous chargez la macro de la fonction "Connect". Cliquez sur la nouvelle commande dans le menu de votre application pour lancer Readiris. L'interface "Connect" peut vous poser quelques questions avant de faire tourner Readiris. Suivez la procédure normale et quittez l'interface "Connect" lorsque c'est fait. Les résultats de reconnaissance sont insérés à l'endroit où se trouve le curseur.

## **OBTENIR DE L'AIDE EN LIGNE**

Ceci termine l'aperçu sur Readiris. Quelques informations de dernière minute peuvent ne pas se trouver dans ce manuel. Nous vous recommandons donc de consulter le système d'aide en ligne pour obtenir des renseignements complémentaires sur Readiris.

Pour ce faire, rendez-vous au menu "Aide". La commande "Rubriques d'Aide" ou son raccourci clavier F1 vous permet de naviguer au travers des nombreux sujets expliqués.

