SmartSort 1.1

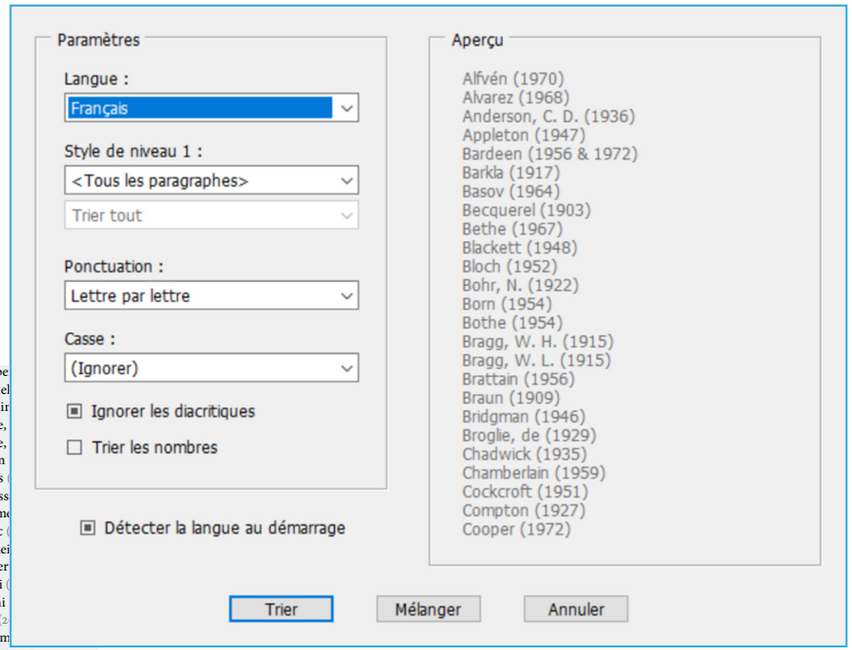
Tri alphabétique avancé dans InDesign

MANUEL D'UTILISATION

Aperçu

1. Description

SmartSort permet de trier alphabétiquement une liste de mots ou d'expressions en respectant les conventions d'une langue donnée. Par défaut, le script agit sur les paragraphes sélectionnés. Un tri de second ordre est également possible si vous spécifiez un style de paragraphe dit « de niveau 1 ». Les termes sont réordonnés par un algorithme robuste pouvant traiter jusqu'à 65 535 items. Des paramètres avancés permettent d'ajuster ce tri selon les signes de ponctuation, la casse, les diacritiques, les chiffres. SmartSort supporte principalement les alphabets latin, grec et cyrillique. Il gère plus de 230 langues, y compris des sous-systèmes particuliers comme l'espagnol traditionnel, les annuaires téléphoniques en allemand, le breton, etc.



Le dialogue de SmartSort offre la même apparence sous Mac OS et Windows. L'interface s'ajuste automatiquement dans les langues suivantes :

- ▶ français
- ▶ anglais *(défaut)*
- ▶ allemand
- ▶ espagnol
- ▶ italien
- ▶ russe

2. Configuration requise

- Mac OS X / macOS (10.6 ou plus récent), *ou* Windows 7 / 8 / 10.
- Fréquence d'horloge CPU d'au moins 3 GHz.
- Mémoire vive (RAM) d'au moins 4 Go.
- Résolution d'écran 800 × 600 minimum.
- Adobe InDesign (CC recommandé).

3. Algorithme open-source

SmartSort intègre un module open-source nommé *Collator*, issu du framework IdExtenso, qui implémente l'*algorithme de collation Unicode* (UCA) de façon simplifiée. Le code source et sa documentation sont disponibles en clair à cette adresse : github.com/indiscripts/IdExtenso/blob/master/etc/$$.Collator.jsxlib

Pour sa part, SmartSort est téléchargeable gratuitement ici : indiscripts.com/category/projects/SmartSort

Prise en main

1. Avant d'installer

SmartSort reside dans un seul fichier : SmartSort.jsx. Toutefois, cet élément est délivré au format ZIP lors de votre téléchargement. La première étape consiste donc à *dé-zipper* et extraire le fichier SmartSort.jsx, qu'il conviendra ensuite de placer dans un dossier spécifique, comme indiqué ci-après.

NOTE Avant de procéder à l'installation, il est recommandé de sauvegarder vos documents de travail et de redémarrer InDesign.

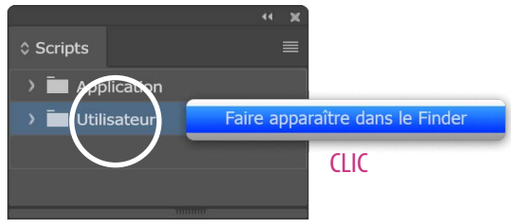
2. Installation sous Mac

1) Depuis InDesign, affichez le panneau Scripts *via* :
   - Fenêtre ▸ Utilitaires ▸ Scripts (CC, CS6, CS5), ou
   - Fenêtre ▸ Automatisation ▸ Scripts (CS4).

MAC OS

2) Faites un CLIC DROIT sur le dossier *Utilisateurs*, puis cliquez sur « Faire apparaître dans le Finder ».

3) Cette procédure révèle un dossier Scripts Panel. Déposez le fichier SmartSort.jsx dans ce dossier.



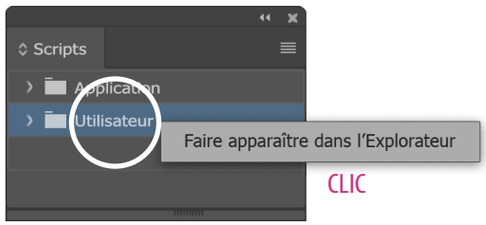
CLIC DROIT

3. Installation sous Windows

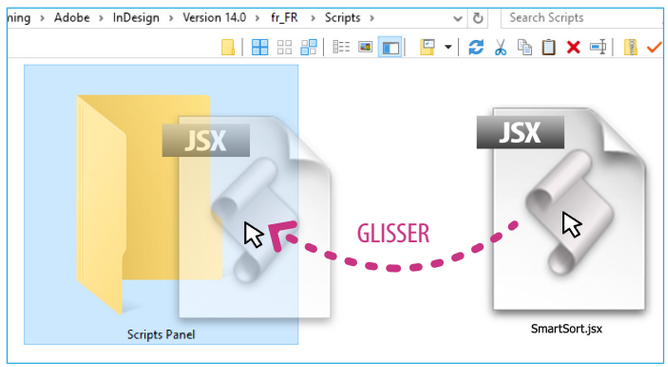
1) Depuis InDesign, affichez le panneau Scripts *via* :
   - Fenêtre ▸ Utilitaires ▸ Scripts (CC, CS6, CS5), ou
   - Fenêtre ▸ Automatisation ▸ Scripts (CS4).

2) Faites un CLIC DROIT sur le dossier *Utilisateurs*, puis cliquez sur « Faire apparaître dans l'Explorateur ».

3) Cette procédure révèle un dossier Scripts Panel. Déposez le fichier SmartSort.jsx dans ce dossier.



CLIC DROIT



WINDOWS

4. Installer une mise à jour

Si vous êtes avisé(e) qu'une mise à jour de SmartSort est disponible, téléchargez le nouveau pack ZIP et procédez comme suit :

1) Quittez InDesign.

2) Après dézippage, déposez le fichier SmartSort.jsx à la place du précédent (donc dans le même dossier et sous le même nom).

3) Redémarrez InDesign.

NOTE La nouvelle version est immédiatement fonctionnelle, vos préférences sont conservées.

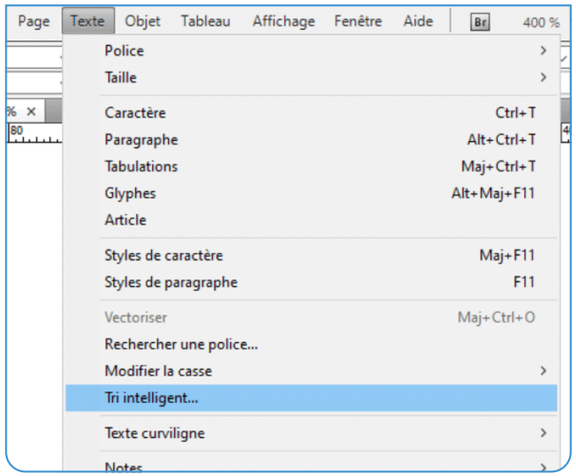
5. Premier démarrage depuis le panneau Scripts

Une fois le script installé, retournez dans InDesign. Le premier appel de SmartSort devra se faire depuis le panneau Scripts :

1) Affichez le panneau Scripts *via* :
   - Fenêtre ▸ Utilitaires ▸ Scripts (CC, CS6, CS5), ou
   - Fenêtre ▸ Automatisation ▸ Scripts (CS4).

2) Dans la branche *Utilisateur*, recherchez SmartSort.jsx puis DOUBLE-CLIQUEZ sur ce dernier.

6. Appel de SmartSort par le menu d'InDesign



SmartSort s'intègre au système de menus d'InDesign, si bien que vous pouvez l'invoquer à tout moment, sans avoir à ouvrir le panneau Scripts.

Une commande dédiée est automatiquement associée à SmartSort dans le menu Texte, sous l'item *Modifier la casse*. L'intitulé apparaît en français (*Tri intelligent…*) à moins que votre version d'InDesign adopte une autre langue :

▸ Français	`Tri intelligent...`
▸ Anglais	`Smart Sort...`
▸ Allemand	`Intelligente Sortierung...`
▸ Espagnol	`Clasificación inteligente...`
▸ Italien	`Ordinamento intelligente...`
▸ Russe	`Умная сортировка...`

Quand un texte est sélectionné, la commande `Tri intelligent...` est également disponible dans le menu contextuel (CLIC DROIT).

7. Test rapide

Vous voici fin prêt(e) à trier une liste de mots à la vitesse de l'éclair :

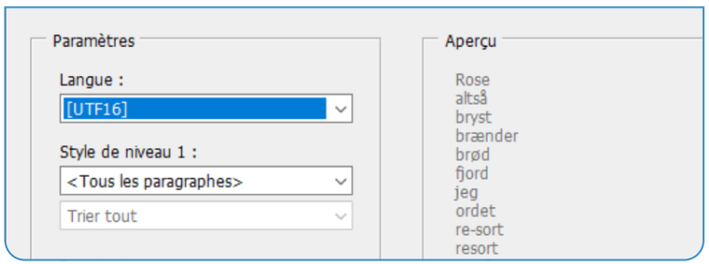
1) Ouvrez ou créer un document InDesign.

2) Copiez dans le presse-papier la liste (non triée) ci-contre.

3) Créez un bloc-texte vide (outil Texte) et collez les lignes.

4) Sélectionnez le bloc et invoquez Texte ▸ *Tri intelligent…*

5) À la rubrique Langue, sélectionnez d'abord l'élément `[UTF16]` en tête de liste. Le panneau Aperçu s'actualise en conséquence, affichant le classement de base UTF-16.



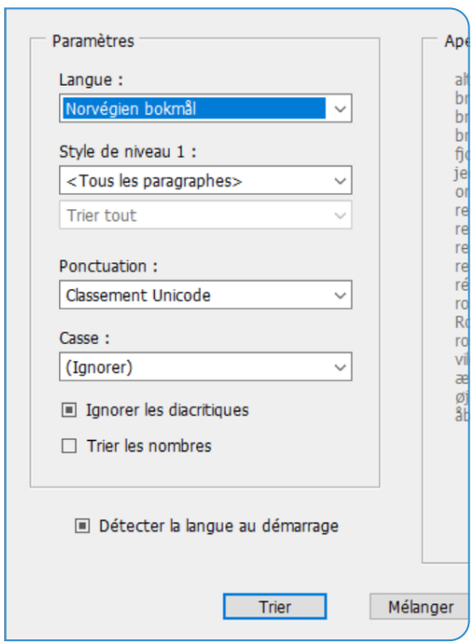
NOTE *Grosso modo*, le système UTF-16 reflète l'ordre intrinsèque des caractères Unicode situés dans les tables de base (notamment l'Ascii, l'alphabet latin, etc.).

SANS TRI	[UTF16]	Europe [EOR]	Bokmål
fjord	Rose	åbner	altså
altså	altså	ære	bryst
resume	bryst	altså	brænder
vild	brænder	brænder	brød
result	brød	brød	fjord
øje	fjord	bryst	jeg
jeg	jeg	fjord	ordet
re-sort	ordet	jeg	re-sort
åbner	re-sort	øje	resort
ordet	resort	ordet	result
brød	result	re-sort	resume
brænder	resume	resort	résumé
rosé	rose	result	Rose
resort	rosé	resume	rose
bryst	résumé	résumé	rosé
résumé	vild	Rose	vild
Rose	åbner	rose	ære
rose	ære	rosé	øje
ære	øje	vild	åbner

6) Sélectionnez à présent le système `Europe [EOR]` au lieu de `[UTF16]`. Notez que l'aperçu propose maintenant un classement plus cohérent des digrammes et des lettres accentuées (å, æ, é, ø).

7) Enfin, choisissez la langue `Norvégien bokmål` et cliquez sur le bouton Trier.

NOTE Dans les langues danoises et norvégiennes, les formes å, æ et ø constituent des lettres distinctes classées après le z.



Paramètres linguistiques

L'opération que l'on appelle communément TRI ALPHABÉTIQUE repose sur des conventions liées au système d'écriture. Bien qu'un ALPHABET donné fixe par lui-même un ordre (entre les lettres), il ne détermine pas à lui seul toutes les règles portant sur l'ordonnancement des mots dans n'importe quelle langue employant cet alphabet. Selon les cultures, dialectes ou conventions locales, des règles additionnelles entrent en jeu.

1. Langues et alphabets supportés

SmartSort reconnaît les lettres et signes de ponctuation de 230 langues. Le script discrimine environ 10 000 caractères et séquences listés dans la *table de comparaison par défaut d'Unicode* (Default Unicode Collation Element Table, en abrégé DUCET). Les tableaux ci-contre donnent un rapide aperçu des principaux blocs pris en charge.

SmartSort gère la plupart des langues basées sur les alphabets latin, grec et cyrillique. D'autres systèmes d'écriture tels que l'hébreu, l'arabe, le devanagari, le thaï ou l'arménien sont pris en charge *a minima*, c'est-à-dire en appliquant l'ordre défini par le DUCET, sans autre ajustement.

NOTE Les idéogrammes unifiés et autres caractères asiatiques ne sont pas reconnus par SmartSort.

L'AJUSTEMENT (*tailoring* en anglais) est un processus permettant d'affiner les règles ou l'algorithme de tri pour certaines langues.

BLOC UNICODE	INTERVALLE
Basic Latin	0000..007F
Latin-1 Supplement	0080..00FF
Latin Extended-A	0100..017F
Latin Extended-B	0180..024F
Latin Extended Additional	1E00..1EFF
Latin Extended-C	2C60..2C7F
Latin Extended-D	A720..A7FF
Latin Extended-E	AB30..AB6F
Greek and Coptic	0370..03FF
Greek Extended	1F00..1FFF
Cyrillic	0400..04FF
Cyrillic Supplement	0500..052F
Cyrillic Extended-C	1C80..1C8F
Cyrillic Extended-A	2DE0..2DFF
Cyrillic Extended-B	A640..A69F
Armenian	0530..058F
Hebrew	0590..05FF
Arabic	0600..06FF
Arabic Supplement	0750..077F
Arabic Extended-A	08A0..08FF
Devanagari	0900..097F
Devanagari Extended	A8E0..A8FF
Bengali	0980..09FF
Tamil	0B80..0BFF
Telugu	0C00..0C7F
Malayalam	0D00..0D7F
Thai	0E00..0E7F
Lao	0E80..0EFF

SmartSort gère essentiellement les alphabets latin, grec et cyrillique. Certains systèmes additionnels sont partiellement pris en charge : arménien, hébreu, arabe…

BLOC UNICODE	INTERVALLE
Alphabetic Presentation Forms	FB00..FB4F
Arabic Mathematical Alphabetic Symbols	1EE00..1EEFF
Arabic Presentation Forms-A	FB50..FDFF
Arabic Presentation Forms-B	FE70..FEFF
Combining Diacritical Marks	0300..036F
Combining Diacritical Marks Extended	1AB0..1AFF
Combining Diacritical Marks Supplement	1DC0..1DFF
Combining Half Marks	FE20..FE2F
Enclosed Alphanumerics	2460..24FF
General Punctuation	2000..206F
Halfwidth and Fullwidth Forms	FF00..FFEF
IPA Extensions	0250..02AF
Letterlike Symbols	2100..214F
Mathematical Alphanumeric Symbols	1D400..1D7FF
Small Form Variants	FE50..FE6F
Spacing Modifier Letters	02B0..02FF
Superscripts and Subscripts	2070..209F
Supplemental Punctuation	2E00..2E7F

De nombreux caractères supplémentaires seront aussi examinés : diacritiques, modificateurs, ponctuation, symboles alphanumériques, etc.

Ce mécanisme intervient dans de nombreux cas examinés par l'algorithme de collation Unicode (UCA). Par exemple, le digramme `ch` est traité en espagnol traditionnel comme une lettre à part entière, classée entre `c` et `d`. Le même phénomène apparaît pour le digramme `ll`. En breton, le digramme `ch` et le trigramme `c'h` sont également considérés comme des lettres autonomes, classées entre `c` et `d`. Inversement, des ligatures usuelles telles que Æ,

IJ, Œ, ß sont ordonnées respectivement comme les digrammes AE, IJ, OE, ss dans la plupart des langues européennes. (Ces règles sont gérées collectivement par le classement dit [EOR], *European Ordering Rules*.)

2. Comparaison multi-niveaux

SmartSort s'appuie sur un mécanisme de comparaison à trois niveaux hérité de l'algorithme UCA :

- *Niveau 1 (lettres fondamentales)* *Ex. :* bas < bus
  Lors de la comparaison de deux expressions, le critère primordial est la position de chaque lettre dans l'ordre alphabétique sans tenir compte des accents ou différences de casse. Il s'agit de la COMPARAISON DE NIVEAU 1. Notez qu'une *lettre*, à ce stade, ne coïncide pas nécessairement avec un caractère Unicode (songez au digramme CH en breton ou au digramme DŽ en serbo-croate). D'autre part, un élément perçu comme une lettre accentuée dans telle langue peut se comporter comme une lettre fondamentale dans telle autre. Ainsi, Ñ s'ordonne après N en espagnol, Å s'ordonne après Z en suédois, etc.

- *Niveau 2 (accents et diacritiques)* *Ex. :* tache << tâche
  Lorsque deux lettres fondamentales sont équivalentes au niveau 1 (comme A et Â dans la plupart des langues latines), le classement s'affine en fonction des accents ou signes diacritiques. Cette COMPARAISON DE NIVEAU 2 est requise, du moins souhaitable, dans certaines langues ou cultures. Par exemple, les annuaires

LANGUE	AJUSTEMENTS *(aperçu)*
[EOR] European Ordering Rules LATN/GREC	æ=ae ; œ=oe…
Afrikaans	n << ŉ
Albanais	c < ç ; d < dh … x < xh ; z < zh
Allemand [annuaire]	ae << ä ; oe << ö…
Allemand [annuaire/Autriche]	a < ä ; o < ö ; u < ü…
Anglais	æ=ae ; œ=oe
Arménien	բ < և <<< ԵՆ
Asturien	*cf. Espagnol [traditionnel]*
Azéri LATN/CYRL	c < ç ; e < ə … и < ы
Basque	*cf. Espagnol [moderne]*
Bas-sorabe	c < č < ć … k < ł…
Biélorusse	e < ë ; y < ў
Bosniaque LATN/CYRL	c < č < ć ; d < dž < đ ; l < lj…
Breton	c < ch < c'h
Catalan	ll << l·l <<< L·l <<< L·L
Cebuano	*cf. Philippin*
Croate LATN/CYRL	*cf. Bosniaque*
Danois	d << đ … z < æ < ø < å…
Espagnol [moderne]	n < ñ
Espagnol [traditionnel]	c < ch ; l < ll ; n < ñ
Espéranto	c < ĉ ; g < ĝ ; h < ĥ…
Estonien	s < š < z < ž ; v < õ < ä < ö…
Éwé	d < dz < ɖ ; e < ɛ ; f < ƒ…
Féroïen	a < á ; d < ð … z < æ < ø…
Finnois	v << w ; y << ü … z < æ < ø…
Frioulan	a < à < á < â ; e < è ; i < ì…
Frison occidental	i << y
Galicien	*cf. Espagnol [traditionnel]*
Gallois	c < ch ; d < dd…
Groenlandais	q << ĸ ; z < æ < ø < å…
Haoussa *(boko)* LATN *TRANSLIT. ARAB*	b < ɓ ; d < ɗ ; k < ƙ ; s < sh…
Haut-sorabe	c < č ; d < dź ; e < ě…
Hawaïen	a < e <<< E < i … w < ʻ
Hébreu	' ~ <GERESH> ; " ~ <GERSHAYIM>
Hongrois	c < cs ; d < dz < dzs ; g < gy…
Igbo	b < ch ; g < gb < gh < gw…
Islandais	a < á ; d < đ … z < þ < æ…
Kannada	೦ < ೧ < ೨ < ೩ < ೪ < ೫ < ೬…
Kazakh	e < ë ; ұ < ү ; ы < i

LANGUE	AJUSTEMENTS *(aperçu)*
Kirghize	e < ë
Kurde	c < ç ; e < ê ; i < î…
Letton	c < č ; g < ģ … z < ž
Lingala	e < ɛ ; o < ɔ
Lingala [phonétique]	e < ɛ ; o < ɔ ; g < gb ; k < kp…
Lituanien	a << ą ; a < č … s < š…
Macédonien	ѓ < ѕ … ћ < ќ
Maltais	b < ċ ; f < ġ ; g < għ … y < ż
Moldave LATN/CYRL	*cf. Roumain*
Néerlandais	ij=ij
Néerlandais [annuaire]	ij=y
Norvégien bokmål	*cf. Danois*
Oromo	z < ch < dh < kh < ny < ph < sh
Ouzbek LATN/CYRL	z < oʻ < gʻ < sh … e < ë
Philippin	n < ñ < ng
Polonais	a < ą ; c < ć … z < ź < ż
Roumain LATN/CYRL	a < ă < â ; i < î ; s < ş ; t < ţ
Sami du Nord	a < á ; c < č … z < ž < æ…
Serbe LATN/CYRL	*cf. Bosniaque*
Slovaque	a < ä ; c < č ; h < ch … z < ž
Slovène	c < č ; s < š ; z < ž
Suédois	z < å < ä < ö << ø
Tatar	я < ә < ө < ү < җ < һ < ь
Tchèque	c < č ; h < ch…
Tonguien	n < ng ; z < ʻ …
Turc	c < ç ; g < ğ ; h < ı ; o < ö…
Turkmène LATN/CYRL	c < ç ; e < ä … e < ë ; ə < ə
Ukrainien	г < ґ ; я < ь…
Vietnamien [standard]	a < ă < â ; d < đ ; e < ê…
Wolof	a < à ; e < é < ë ; n < ñ < ŋ…
Yoruba	e < ẹ ; g < gb ; o < ọ ; s < ṣ
Yoruba [Bénin]	e < ɛ ; g < gb ; k < kp ; o < ɔ…

Langues faisant l'objet d'un ajustement spécifique depuis SmartSort 1.13 et aperçu des règles mises en œuvre.
Écriture latine · Écriture cyrillique
Alphabets mixtes · Autres systèmes

téléphoniques allemands traitent le caractère ä comme une variante du digramme ae mais le classent après celui-ci. (Ce qu'on exprime par la règle ae << ä.) La même chose se produit en catalan avec les formes ll et l·l.

► *Niveau 3 (casse)* *Ex. :* `pierre <<< Pierre`
Les différences de casse (majuscules *vs* minuscules) sont habituellement transparentes, *sauf lorsque les lettres et accents sont rigoureusement identiques.* Une COMPARAISON DE NIVEAU 3 peut alors s'engager si besoin, selon les spécifications de l'éditeur ou les options particulières de l'ouvrage en cours.

NOTE Gardez à l'esprit que la comparaison de niveau 3 affecte uniquement l'ordre des termes *déjà équivalents aux niveaux 1 et 2.* Ainsi, l'option « Minuscules en premier » produira l'ordre `pierre <<< Pierre < pizza <<< Pizza < pot`, alors que « Majuscules en premier » donnerait `Pierre <<< pierre < Pizza <<< pizza < pot`.

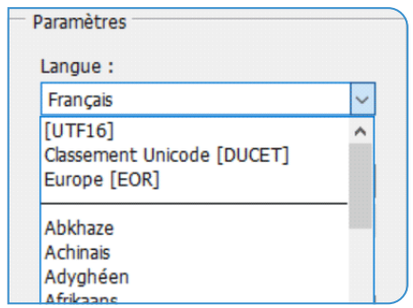
3. Choix de la langue

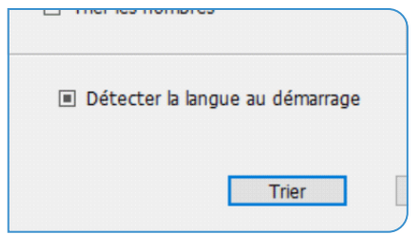
1) Sélectionnez dans InDesign les paragraphes à ordonner.

2) Exécutez SmartSort depuis le menu Texte ► *Tri intelligent…*

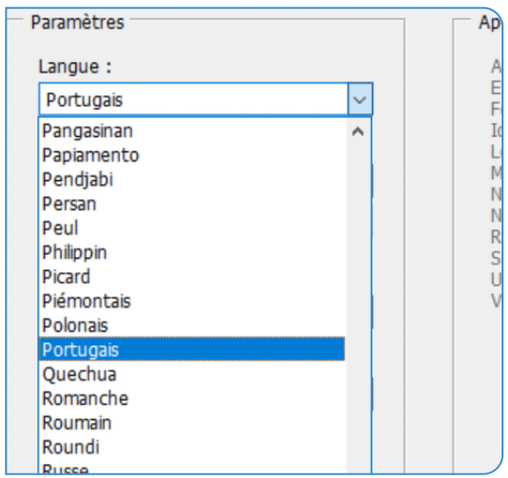
3) Dans la liste déroulante *Langue*, cliquez sur l'élément désiré parmi les 230 choix disponibles. Certains ajustements avancés sont proposés ici, tels que *francoprovençal*, *allemand (annuaire)*, *espagnol (traditionnel)*, etc. Si aucune langue ne convient, choisissez l'une des options suivantes :

- `Europe [EOR]` si les caractères employés reflètent essentiellement les alphabets latin, grec et/ou cyrillique ;
- `Classement Unicode [DUCET]` si des caractères exotiques interviennent mais tolèrent le classement par défaut (DUCET) ;
- `[UTF16]` pour les données alphanumériques n'impliquant aucune langue particulière (références de produits, chaînes Ascii, etc).





NOTE Si l'option « Détecter la langue au démarrage » est activée, le script sélectionne automatiquement la langue qui lui paraît la plus appropriée selon le texte sélectionné.

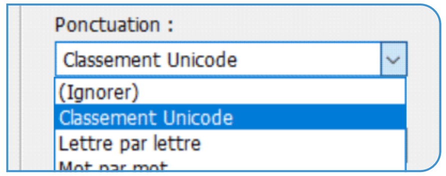


4. Ponctuation

Espaces, traits d'union, virgules et quelques centaines de signes analogues sont identifiés comme « éléments à poids variable » (*variable collation elements*). Contrairement aux caractères non répertoriés que l'algorithme ignore purement et simplement, les éléments à poids variable peuvent influencer le classement. Le répertoire DUCET d'Unicode leur attribue des valeurs spéciales conduisant à l'arrangement suivant :

1. Espaces.
2. Traits d'union et assimilés.
3. Virgule, point-virgule, deux-points.
4. Points d'exclamation et d'interrogation.
5. Point(s), points de suspension.
6. Guillemets et apostrophes.
7. Parenthèses, crochets, accolades.
8. Autres symboles.

La rubrique Ponctuation propose alors quatre choix : *(Ignorer)*, *Classement Unicode*, *Lettre par lettre*, *Mot par mot.*



- *(Ignorer)*
Sélectionnez cette option pour ignorer l'effet des éléments à poids variable. Le tri est donc aveugle aux signes de ponctuation et considère les formes `In Design`, `in-design`, `(InDesign)`, `#indesign`… comme purement équivalentes — sauf à prendre en compte les différences de casse.

- *Classement Unicode*
Cette option applique aux éléments variables le poids que leur assigne le DUCET. Ainsi, les espaces se rangent avant les traits d'union, ces derniers avant les parenthèses, etc., comme illustré ci-dessous.

(Ignorer)	Classement Unicode
Illustrator	_photoshop
InDesign	(InDesign)
In Design	[Photoshop]
in-design	#indesign
(InDesign)	Illustrator
#indesign	In Design
InDesign, by Adobe	in design we trust
InDesign (CC)	in-design
InDesign CS	in.design.we.trust
InDesign & PS	InDesign
in design we trust	InDesign (CC)
in.design.we.trust	InDesign & PS
Kuler	InDesign CS
Photoshop	InDesign, by Adobe
photo shop	Kuler
_photoshop	photo shop
[Photoshop]	Photoshop

Effet du *Classement Unicode* des signes de ponctuation.
• À gauche, les blocs surlignés indiquent des termes équivalents.
• À droite, les caractère surlignés influent sur le classement.

- » TABULATION
- · ESPACE
- ^ INSÉCABLE
- CADRAT
- FLUSH SPACE
- DEMI-CADRATIN
- CADRATIN
- TIERS D'ESPACE
- QUART D'ESPACE
- SIXIÈME D'ESPACE
- # ESPACE TABULAIRE
- ! PONCTUATION
- ESPACE FINE
- ULTRAFINE
- ESPACE MATH.
- ‾ TIRET HAUT
- _ TIRET BAS
- ‗ DOUBLE TIRET BAS
- - MOINS
- - TRAIT D'UNION
- ‒ TIRET NUMÉRIQUE
- – DEMI-TIRET
- — TIRET CADRATIN
- ― BARRE HORIZ.
- , ، ، ، VIRGULES dont virgules arménienne et arabe, séparateurs décimaux, etc.
- ; POINT-VIRGULE
- ؛ PT-VIRGULE ARABE
- : DEUX-POINTS
- ! ‼ ¡ ? ¿ ‽ ⸘ POINTS D'EXCLAMATION, D'INTERROGATION ET VARIANTES
- . POINT
- . POINT DE SUITE
- .. DEUX POINTS
- … P. DE SUSPENSION
- · POINT MÉDIAN
- ' ' ‚ ‛ ‹ › " " „ ‟ « » GUILLEMETS ET APOSTROPHES
- ( ) [ ] { } PARENTHÈSES, CROCHETS, ACCOLADES
- § SECTION
- ¶ ALINÉA
- @ AROBASE
- * ASTÉRISQUE
- / BARRE OBLIQUE
- \ B. OBLIQUE INV.
- & ESPERLUETTE
- # SYMBOLE NUM.
- % POUR CENT
- ‰ POUR MILLE
- † DAGUE
- ‡ DAGUE DOUBLE
- ° DEGRÉ
- ҂ MILLIERS (CYRIL.)
- + PLUS
- ± PLUS OU MOINS
- ÷ DIVISION
- × MULTIPLICATION
- < INFÉRIEUR
- = ÉGAL
- > SUPÉRIEUR
- | LIGNE VERTICALE
- ¦ LIGNE BRISÉE
- ~ TILDE

Quelques éléments variables rangés par poids croissant. Cet ordre s'applique quand l'option *Classement Unicode* est retenue.

Paramètres linguistiques

- *Lettre par lettre*

  Le classement dit « lettre par lettre » ignore les signes de ponctuation *sauf les parenthèses et les virgules* employées comme séparateurs. Ce système a prédominé aux États-Unis, surtout dans l'édition encyclopédique. Le *Chicago Manual of Style*, guide orthotypographique de référence outre-Atlantique, définit les règles de précédence comme suit : mot, mot suivi d'une parenthèse, mot suivi d'une virgule, puis — abstraction faite des espaces et autres signes de ponctuation — mot suivi d'un nombre et mot suivi d'autres lettres.

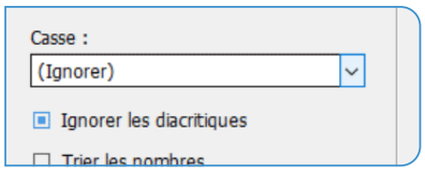
  Cette option est semblable à *(Ignorer)* en ce qu'elle inhibe les espaces et la plupart des signes de ponctuation — notamment les traits d'union des mots composés. La différence cruciale est qu'une parenthèse ou une virgule interrompt le processus et positionne le terme analysé avant ses formes concurrentes.

- *Mot par mot*

  Le classement dit « mot par mot » ignore les signes de ponctuation *sauf les parenthèses, virgules, espaces et traits d'union*. Ce système est préféré au Royaume-Uni et recommandé par plusieurs normes internationales. Selon l'*Oxford Guide to Style*, guide orthotypographique de référence outre-Manche, les éléments séparés par un trait d'union dans un mot composé devraient être regardés comme des mots distincts, à moins que le premier élément ne constitue pas un véritable mot. Cette subtilité ne pouvant être tranchée par l'algorithme de tri, SmartSort considère ici que tout trait d'union vaut pour une espace mot. Dès lors, les espaces comme les traits d'union sont classés après les parenthèses et les virgules, dans cet ordre : A(B < A,B < A B. Par ailleurs, espaces et traits d'union sont ignorés lorsque suivis d'une parenthèse ouvrante ou d'une virgule, de sorte que la forme A (B équivaut à A(B et se classe avant A,B.

(Ignorer)	Lettre par lettre	Mot par mot
New, Arthur	NEW (Neighbors Ever Watchful)	NEW (Neighbors Ever Watchful)
newborn	NEW (Now End War)	NEW (Now End War)
newcomer	New, Arthur	New, Arthur
New Deal	New, Zoe	New, Zoe
new economics	newborn	New Deal
newel	newcomer	new economics
New England	New Deal	New England
“new-fangled notions”	new economics	“new-fangled notions”
Newfoundland	newel	new math
newlyweds	New England	New Thorndale
new math	“new-fangled notions”	new town
NEW (Neighbors Ever Watchful)	Newfoundland	New Year’s Day
NEW (Now End War)	newlyweds	newborn
new/old continuum	new math	newcomer
newsboy	new/old continuum	newel
news conference	news, lamentable	Newfoundland
news, lamentable	*News,Networks&Arts*	newlyweds
newsletter	newsboy	new/old continuum
News,Networks&Arts	news conference	news, lamentable
News of the World (Queen)	newsletter	*News,Networks&Arts*
news release	*News of the World* (Queen)	news conference
newt	news release	*News of the World* (Queen)
New Thorndale	newt	news release
NEWT (North. Estuary Wind Tunnel)	NEWT (North. Estuary Wind Tunnel)	newsboy
new town	New Thorndale	newsletter
New Year’s Day	new town	newt
New, Zoe	New Year’s Day	NEWT (North. Estuary Wind Tunnel)

Tris LETTRE PAR LETTRE et MOT PAR MOT (en anglais). Le premier système ignore la ponctuation sauf (and , alors que le second traite aussi les espaces et les traits d'union comme séparateurs, ce qui tend à réunir les mêmes mots clés.



5. Casse et signes diacritiques

Les options de tri sensibles à la casse ou aux diacritiques agissent seulement sur l'ordre des expressions *équivalentes au niveau 1*. Il s'agit donc d'un réglage fin ne devenant perceptible que si les lettres fondamentales n'ont pas suffi à différencier et interclasser deux termes de la liste.

NOTE Toutefois, dans certaines langues, une modification diacritique peut se comporter comme une lettre fondamentale. Par exemple, les formes CARE et ÇARË sont parfaitement distinctes en albanais, où C < Ç.

▸ *Casse*

— *Majuscules en premier* range les variantes majuscules avant les autres formes : `adios` < `A.D.O` <<< Ado <<< ado < Adobe.

— *Minuscules en premier*, au contraire, promeut d'abord les variantes minuscules : `adios` < ado <<< Ado <<< `A.D.O` < Adobe.

— *(Ignorer)* ne tient pas compte des différences de niveau 3.

▸ *Diacritiques*

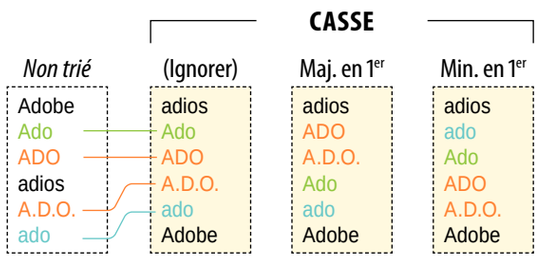
Cochez la case « Ignorer les diacritiques » pour rendre inopérantes les différences de niveau 2. En pratique, cela permet de conserver aux formes concurrentes (par ex., ÈRES et ÉRÈS) leur position relative. Mais la plupart des dictionnaires et index adoptent un classement de niveau 2 en phase avec la langue. Voici quelques-uns des ajustements pris en charge par SmartSort :

– Le trie européen [EOR] traite ə (resp. ɪ, ĸ, ŋ, ŧ) comme variante diacritique de e (resp. i, k, n, t). De plus, þ est considéré comme une variante de th : th << þ.

– En catalan, ll << l·l (ce qui reflète au demeurant l'interclassement [EOR], mais contraste avec l'espagnol traditionnel).

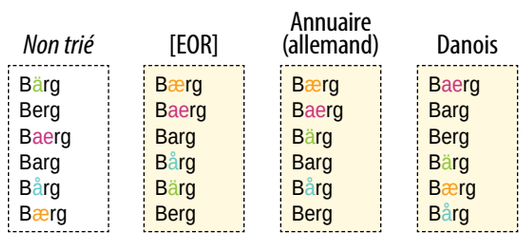
– En finnois, w est une lettre « étrangère » considérée comme variante de niveau 2 de v : v << w.

– Dans les annuaires téléphoniques composés en allemand, ä (resp. ö et ü) est une variante de niveau 2 de ae (resp. oe et ue).



La rubrique « Casse » décide comment interclasser des termes équivalents tels que ADO et ado. Si vous choisissez l'option *(Ignorer)*, ces variantes conserveront leur position relative car SmartSort effectue un tri stable. Si l'option *Majuscules en premier* est activée, les formes ADO et A.D.O.* se rangeront avant Ado et ce dernier avant ado. Ce classement se renverse si vous optez pour *Minuscules en premier*.

* ADO et A.D.O. demeurent équivalents, à moins que le classement Unicode opère sur les signes de ponctuation.



LE TRI DE NIVEAU 2 gère les accents, diacritiques et formes assimilées : ae *vs* ä dans les annuaires allemands, ä *vs* æ en danois, etc.

6. Chiffres et nombres

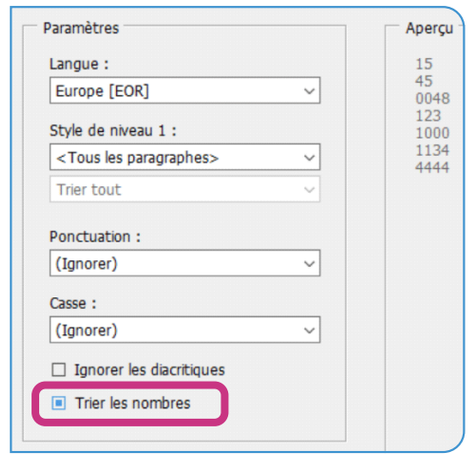
Par défaut, les caractères représentant des chiffres sont ordonnés selon la valeur que leur assigne le DUCET. Ils se classent *après* les signes de ponctuation et *avant* les lettres. Beaucoup d'éléments Unicode équivalents, tels que 1, ¹ (U+00B9), ١ (U+0661), १ (U+0967), ₁ (U+2081), possèdent le même poids fondamental et sont traités en conséquence — à moins que vous optiez pour le mode de classement [UTF16].

Ignorant ces variantes de présentation et autres chiffres exotiques, la fonction *Trier les nombres* concerne exclusivement les caractères 0123456789. Elle détecte les SÉQUENCES DE CHIFFRES formant des nombres isolés, comme dans « *2001, l'Odyssée de l'espace* » (-> 2001) ou « InDesign CC 2022 » (-> 2022). Lesdites séquences sont reconnues à condition qu'elles se détachent de l'expression par une ponctuation quelconque (dite *frontière de mot*) ; ainsi « *ABC123* » ne produirait aucune séquence valide.

NOTE À l'instar d'autres paramètres linguistiques, l'option *Trier les nombres* reste inopérante si le tri [UTF16] est appliqué. Tous les caractères sont alors ordonnés selon leur rang Unicode, quelquefois pertinent, mais incapable de refléter le classement numérique :

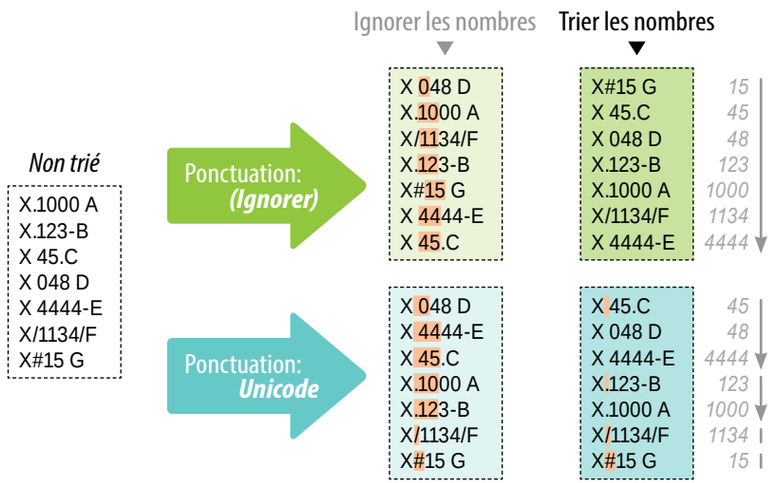
Non trié	[UTF16]
1000	0048
123	1000
45	1134
0048	123
4444	15
1134	4444
15	45

Non trié	Trier les nombres
1000	15
123	45
45	0048
0048	123
4444	1000
1134	1134
15	4444



- *Tri de nombres simples*
  Si votre liste ne comporte que des chiffres (*v.* ci-dessus), choisissez n'importe quelle langue, par ex. [EOR], et cochez la case *Trier les nombres*. Les séquences sont alors interprétées et triées comme des valeurs numériques.

- *Tri d'expressions complexes*
  Si les chiffres se combinent avec d'autres caractères, l'effet de l'option *Trier les nombres* dépend du traitement des signes de ponctuation. Différents réglages sont illustrés ci-dessous :



Trier les nombres produit en général le résultat escompté tant que la ponctuation est inhibée.

Si la ponctuation affecte le classement, chaque caractère significatif (, ., /, #) crée un nouveau groupe de nombres.

Traitement des listes

SmartSort a été conçu pour ordonner rapidement une LISTE DE PARAGRAPHES tout en préservant les styles internes et autres attributs textuels. Dans sa version actuelle, le script peut également manipuler plusieurs blocs-textes d'une traite, ainsi que des lignes de tableau. Enfin, il offre des FONCTIONS DE SOUS-TRI basées sur les niveaux de style à l'échelle du paragraphe.

1. Zone d'aperçu

La zone *Aperçu* de SmartSort permet de visualiser immédiatement l'effet de tel ou tel paramètre avant d'opérer effectivement le tri par le bouton Trier.

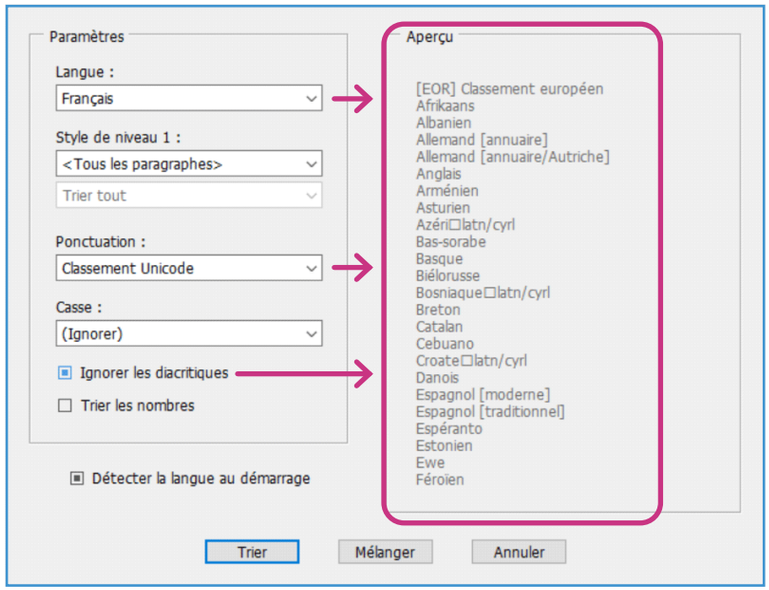
Si le texte sélectionné dans InDesign implique un grand nombre de paragraphes, un échantillon d'une centaine de lignes est chargé en mémoire et l'aperçu affiche les éléments classés en tête (environ une vingtaine de lignes). L'aperçu s'actualise dès qu'une option est modifiée.

2. Sélection des paragraphes cibles

Les PARAGRAPHES SÉLECTIONNÉS déterminent les données entrantes, chaque paragraphe représentant une entrée distincte.

1) Ouvrez votre document InDesign.

2) Placez le curseur dans le bloc-texte ou la cellule cible.



La zone d'aperçu affiche les premiers éléments de la liste triée. Certains caractères spéciaux (tels que la tabulation de retrait à droite) sont signalés par un carré.

3) Sélectionnez les paragraphes à trier. Au moins deux paragraphes sont requis, c'est-à-dire que l'intervalle sélectionné doit comporter au moins un caractère de saut de paragraphe ¶.

NOTE Un paragraphe même partiellement sélectionné est chargé dans son intégralité *(v. ci-contre.)*

NOTE Les sauts de ligne forcés (¬) ne sont pas analysés comme des séparateurs d'entrées. Si votre liste est formatée ainsi, utilisez Rechercher/Remplacer pour convertir ^n en ^p.

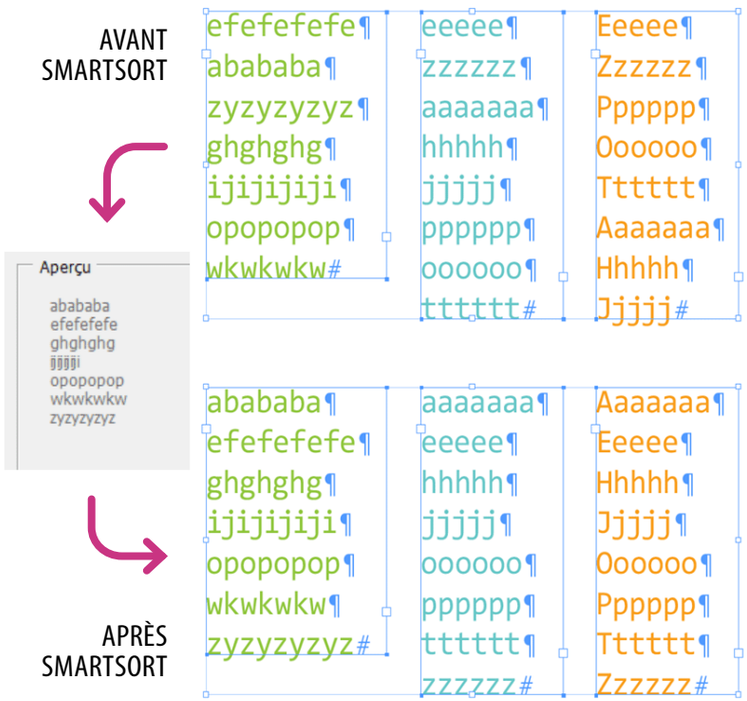
4) Exécutez SmartSort par Texte ▸ *Tri intelligent…* — ou depuis le menu contextuel CLIC DROIT ▸ *Tri intelligent…*

Sélection (InDesign)
Paragraphes cibles (SmartSort)

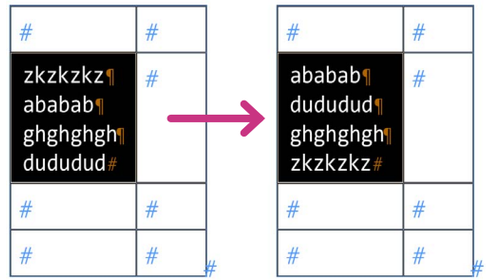
3. Sélection de blocs ou de cellules

▸ *Bloc-texte simple* → *article*
Si vous sélectionnez un bloc-texte par l'outil Sélection, la totalité de l'article est considérée (dont le texte en excès).

▸ *Plusieurs blocs-texte* → *articles (séparés)*
Si vous sélectionnez plusieurs blocs-texte, tous les articles sous-jacents seront pris en charge (texte en excès inclus) et triés séparément. Notez que la zone d'aperçu reflète alors seulement le contenu du premier bloc sélectionné :



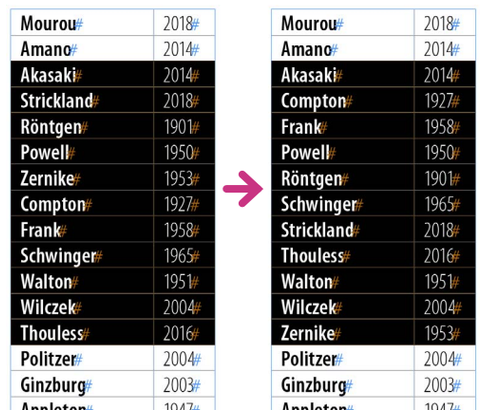
▸ *Cellule(s)* → *paragraphes internes*
— Si une cellule de tableau est sélectionnée (outil Texte), tous ses paragraphes sont considérés (texte en excès inclus).
— Si plusieurs cellules sont sélectionnées, chacune est en principe analysée comme le serait un article autonome. Les paragraphes de chaque cellule sont donc triés. Ce système ne s'applique toutefois que *si le nombre moyen de paragraphes par cellule est supérieur à un (v. ci-dessous.)*



NOTE Dans toutes les situations examinées ici, SmartSort peut s'exécuter aussi bien par Texte > Tri intelligent… que par menu contextuel.

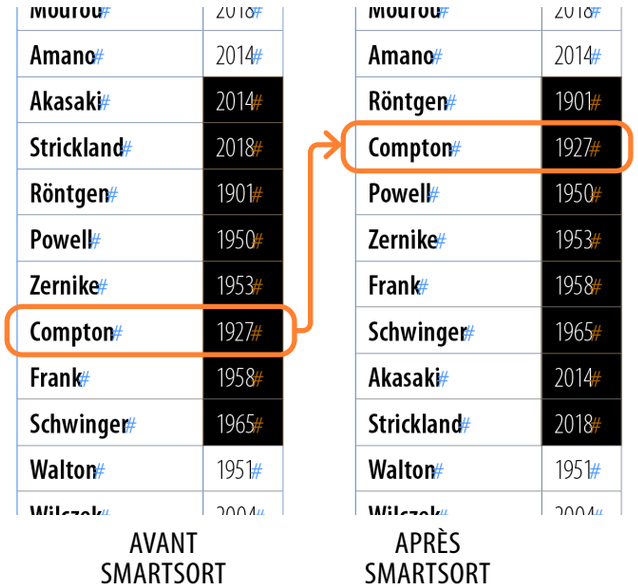
4. Tri de lignes (tableau)

Cas particulier : si vous sélectionnez dans un tableau plusieurs lignes « simples », c'est-à-dire constituées en moyenne d'un seul paragraphe, celles-ci seront triées ensemble et réordonnées les unes par rapport aux autres. ***Cette fonction expérimentale est beaucoup plus lente*** que le tri de paragraphes et ne peut opérer que sur des tableaux simples *sans cellules fusionnées ou scindées.*



Les lignes sont triées ensemble, selon la clé identifiée dans la première colonne.

Les lignes sont triées selon la clé extraite de la première colonne sélectionnée, si bien qu'il est possible de contrôler la clé de tri en sélectionnant une autre colonne :



AVANT SMARTSORT		APRÈS SMARTSORT	
Amano#	2014#	Amano#	2014#
Akasaki#	2014#	Röntgen#	1901#
Strickland#	2018#	Compton#	1927#
Röntgen#	1901#	Powell#	1950#
Powell#	1950#	Zernike#	1953#
Zernike#	1953#	Frank#	1958#
Compton#	1927#	Schwinger#	1965#
Frank#	1958#	Akasaki#	2014#
Schwinger#	1965#	Strickland#	2018#
Walton#	1951#	Walton#	1951#

Dans l'exemple ci-dessus, les lignes sont triées selon la date (2e colonne) et non plus selon le nom.

NOTE Retenez que les opérations dans un tableau induisent un temps d'exécution considérable pour InDesign. Si des centaines de cellules sont visées, prévoyez des dizaines de secondes ou même plusieurs minutes. . .

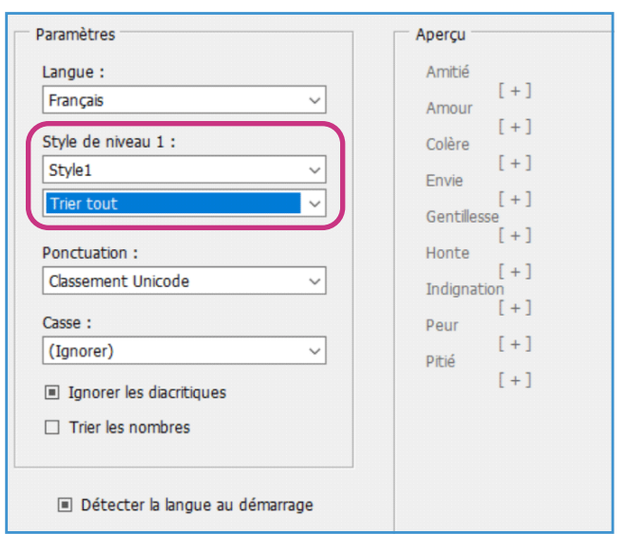
NOTE Si un tableau est entièrement sélectionné, toutes ses lignes de corps sont considérées.

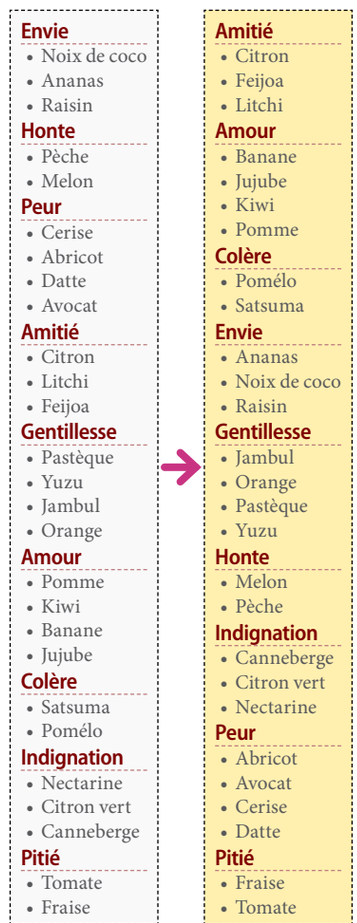
5. Tri et sous-tri

Un document structuré par des STYLES DE PARAGRAPHES peut faire apparaître des niveaux de titres, et donc des regroupements implicites. SmartSort permet de gérer des listes de cette nature sur deux niveaux, pour autant qu'un style de paragraphe soit spécifié comme «Style de niveau 1». Trois possibilités vous sont alors offertes :

▸ *Trier tout*
Réordonne entre eux les paragraphes de niveau 1, ainsi que les éléments de niveau inférieur.

NOTE L'aperçu reflète alors le tri de niveau 1, puis représente chaque sous-liste (triée) par le symbole [+].



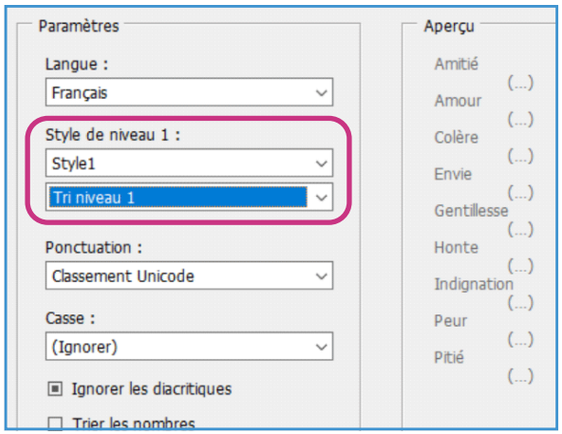


Trier tout

▸ *Tri niveau 1*

Réordonne entre eux les paragraphes de niveau 1 sans affecter l'ordre des paragraphes de niveau inférieur.

NOTE L'aperçu reflète alors le tri de niveau 1 et représente par le symbole (…) les éléments inchangés au niveau inférieur.

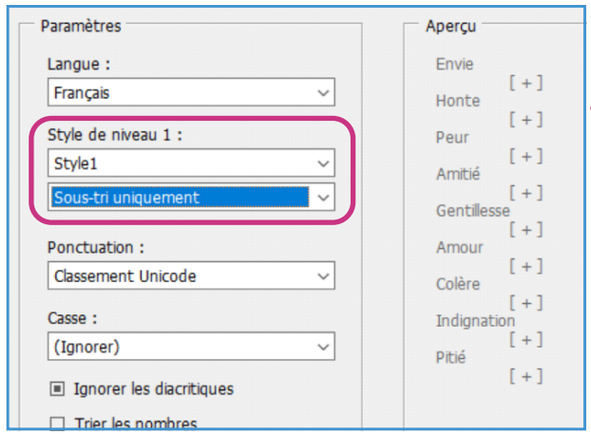


Cette option est bien adaptée au classement des entrées d'un dictionnaire ou d'une encyclopédie, pourvu que chaque mot vedette possède un style dédié.

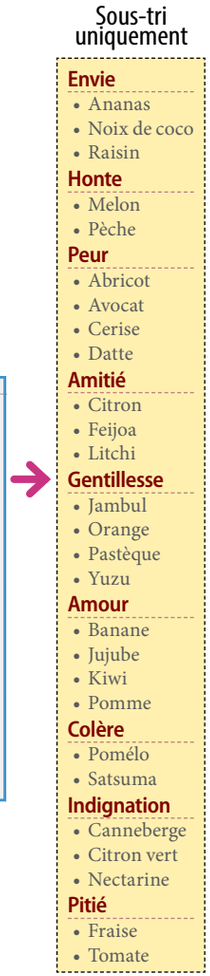
▸ *Sous-tri uniquement*

Conserve l'arrangement des paragraphes de niveau 1, mais réordonne les éléments de niveau inférieur.

NOTE L'aperçu affiche alors les termes de niveau 1 (non triés) et représente chaque sous-groupe (trié) par le symbole [+].

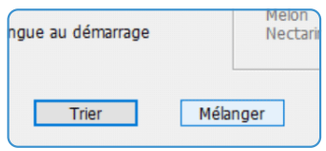


NOTE Il est toujours possible de revenir à un tri « à plat » (sans sous-niveau) en choisissant pour *Style de niveau 1* l'option <Tous les paragraphes>.



6. Mélange

Plutôt que de trier une liste, vous pouvez redistribuer ses éléments au hasard en cliquant sur le bouton Mélanger.



Apostille

SmartSort est un script gratuit et le restera. La mise au point de son algorithme, au plus près des recommandations Unicode, a demandé plusieurs années de travail. La version actuelle est suffisamment *légère* pour opérer dans le moteur de scripts d'InDesign. Elle vous permettra de trier efficacement des listes dans la plupart des contextes.

L'objet principal de ce script étant le tri de paragraphes, il ne peut pas embrasser *toutes* les structures de données comme les tableaux complexes ou les index multi-niveaux. De plus, malgré le nombre élevé de langues prises en charge, il existe évidemment des situations linguistiques récalcitrantes qui demanderaient des ajustements plus fins, des règles *ad hoc*, etc.

Récemment mise en œuvre, la gestion des tableaux « simples » subit certaines limitations techniques, inhérentes à InDesign, pesant lourdement sur le temps d'exécution. Il y a fort à parier que des implémentations plus ciblées et/ou plus efficientes pourraient être développées dans ce cadre particulier.

Ces réserves faites, SmartSort propose une solution solide et originale à un problème récurrent : *le tri alphabétique de données déjà stylées sous InDesign, conformément à une langue de travail.* Ce script reste pourtant méconnu. C'est pourquoi nous vous invitons à le tester par vous-même et à nous adresser sans hésitation vos remarques, commentaires et rapports d'utilisation.

Vous pouvez également soutenir le développement de nos solutions d'automatisation pour InDesign en visitant le site indiscripts.com.

SmartSort 1.1

www.indiscripts.com

Plug-in pour Adobe® InDesign® basé sur Adobe® ExtendScript / ScriptUI. Créé et développé par Marc Autret. Localisation en français, anglais, allemand, espagnol, italien et russe.

Mes remerciements à Peter Kahrel et à Jean-Claude Tremblay, qui m'ont grandement aidé à tester, déboguer et faire évoluer ce projet.

Page d'accueil :	http://indiscripts.com/category/projects/SmartSort
Téléchargement :	http://indiscripts.com/blog/public/scripts/SmartSort.zip
Bugs et feedback :	support@indiscripts.com
Licence d'utilisation :	http://indiscripts.com/pages/eula
Informations de copyright :	http://indiscripts.com/pages/copyright

Ce manuel, tout comme le logiciel qu'il documente, est protégé au titre du droit d'auteur et ne peut être utilisé ou reproduit qu'en accord avec les termes de la licence d'utilisation du produit. Le contenu de ce document est susceptible de corrections ou de modifications sans préavis. Tous les efforts ont été faits pour assurer que les informations de ce document sont correctes. Cependant, Indiscripts ne prend aucun engagement juridique à cet égard. InDesign, les logos InDesign, sont des marques de Adobe Systems Inc.

© Indiscripts, 2021.

Tous droits réservés. Réalisé en France.