SmartSort 1.1

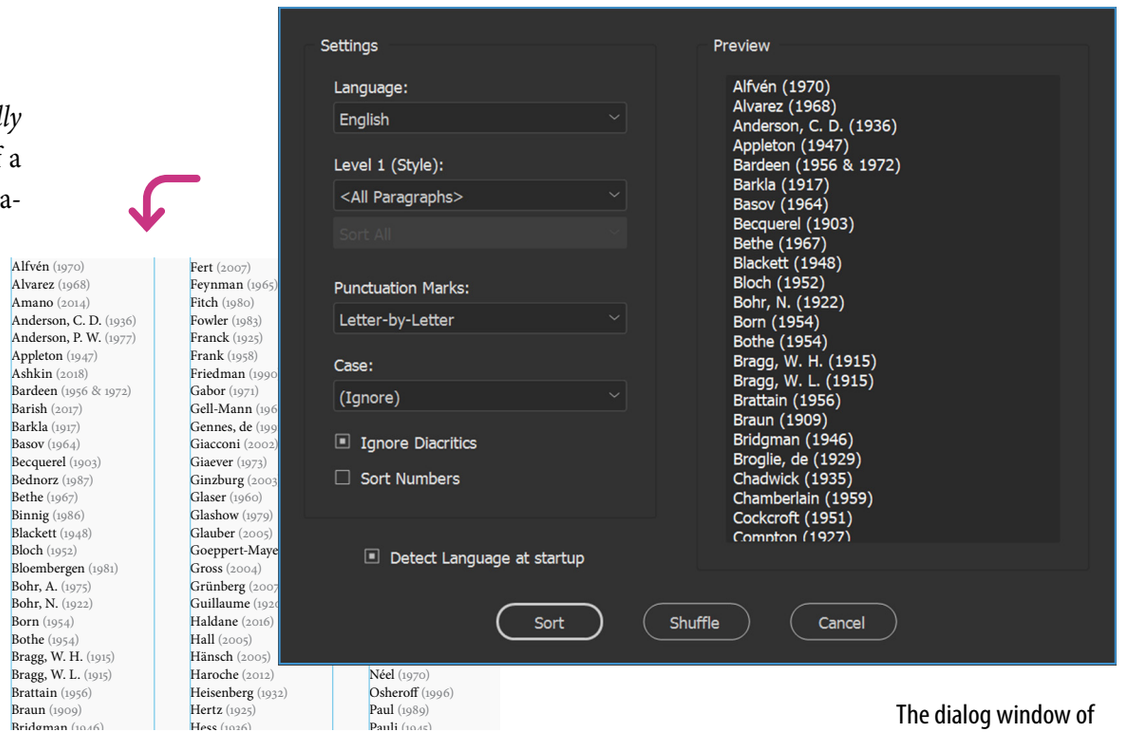
Advanced Paragraph Sorter for InDesign

USER MANUAL

Overview

1. Description

SmartSort brings to InDesign the ability to *sort alphabetically* a set of paragraphs (words, lines) with respect to the rules of a particular language. It basically operates on the selected paragraphs. A *subsort* option is available, based on a primary paragraph style. When you press the **Sort** button, the target paragraphs are reordered using an efficient algorithm that supports up to 65,535 paragraphs. Additional options allow you to fine-tune, apply or ignore advanced sorting rules regarding punctuation marks, case, diacritics, numerals. SmartSort mainly supports Latin, Greek, and Cyrillic alphabets but can address 230+ languages (from Abkhazian to Zulu), including special ordering conventions like in German phonebook, traditional Spanish, Breton…



The dialog window of SmartSort has a very similar look-and-feel in both Mac OS and Windows environments. The following UI locales are available:

- ▶ English *(default)*
- ▶ French
- ▶ German
- ▶ Spanish
- ▶ Italian
- ▶ Russian

2. System requirements

- Mac OS X / macOS (10.6 or later), *or* Windows 7 / 8 / 10.
- CPU with a minimum clock rate of 3 GHz.
- Main memory (RAM) of at least 4GB.
- 800×600 pixel screen-resolution or greater.
- Adobe InDesign CC, CS6, CS5, or CS4.

3. Collation algorithm (open-source)

SmartSort integrates the Collator module, an open-source component from the IdExtenso framework. Collator implements a light-weight version of the Unicode Collation Algorithm (UCA). The source code (incl. full documentation) is available at github.com/indiscripts/IdExtenso/blob/master/etc/$$.Collator.jsxlib

As for SmartSort itself, you can download it for free at indiscripts.com/category/projects/SmartSort

Getting started

1. Before you install

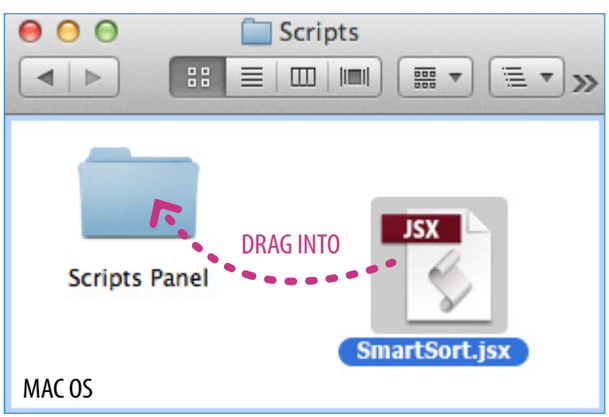
SmartSort resides in a single file: SmartSort.jsx. When you download the script, however, it is originally zipped.

The first step is to unzip the .ZIP file so you can place SmartSort.jsx at the desired location (see below).

As a precaution before you go on, save your working files and restart InDesign in a clean session.

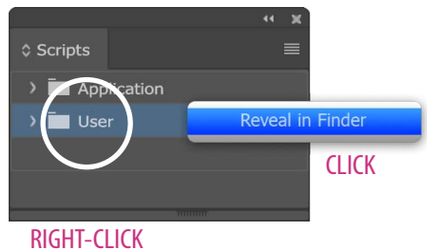
2. Installing in Mac OS

1) In InDesign, open the Scripts panel as follows:
   - Window ▸ Automation ▸ Scripts (CS4), or
   - Window ▸ Utilities ▸ Scripts (CS5, CS5.5, CS6, *or* CC).



2) You see there two main folders: Application and User. Right-click the User folder and pick "Reveal in Finder."

3) You should now see a Scripts Panel folder. Drag SmartSort.jsx into there.

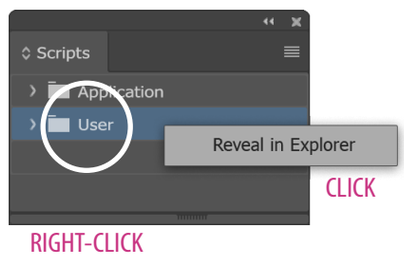


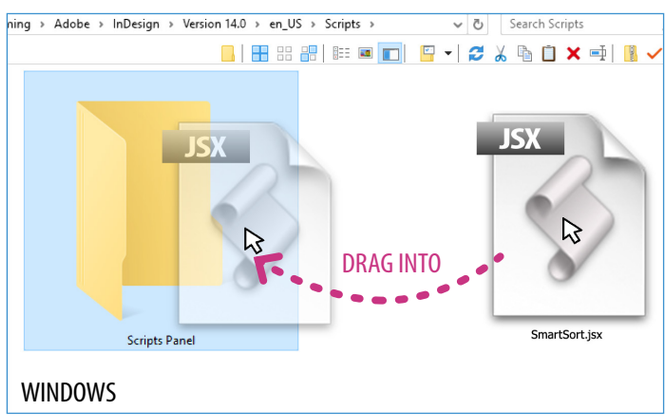
3. Installing in Windows

1) In InDesign, open the Scripts panel as follows:
   - Window ▸ Automation ▸ Scripts (CS4), or
   - Window ▸ Utilities ▸ Scripts (CS5, CS5.5, CS6, *or* CC).

2) You see there two main folders: Application and User. Right-click the User folder and pick "Reveal in Explorer."

3) You should now see a Scripts Panel folder. Drag SmartSort.jsx into there.





4. Installing an update

If you are notified that an update is available, simply download the new package, then

1) Quit InDesign.

2) Unzip and install the file SmartSort.jsx over the previous one, i.e. at the same location.

3) Restart InDesign.

NOTE The new version is instantly functional, and your global settings and preferences are all preserved.

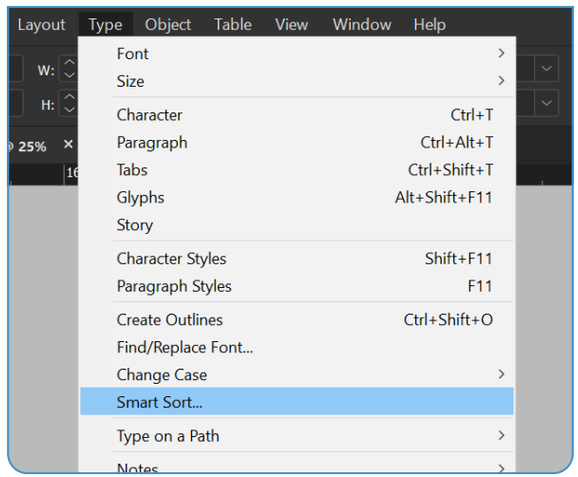
5. First run from the Scripts panel

Once the installation is done, switch back to InDesign. The first execution of SmartSort must be processed from the Scripts panel:

1) In InDesign, display the Scripts panel via:
   - Window ▸ Automation ▸ Scripts (CS4), or
   - Window ▸ Utilities ▸ Scripts (CS5, CS5.5, CS6, *or* CC).

2) Look for SmartSort.jsx in the User folder, then double-click it.

6. Invoking SmartSort from the menu



SmartSort is integrated to InDesign's Type menu, so it can be easily invoked without opening the Scripts panel.

A dedicated command is automatically associated to SmartSort in the Type menu (below the Change Case element.) Its title depends on your InDesign locale:

▸ English *(default)*	`Smart Sort...`
▸ French	`Tri intelligent...`
▸ German	`Intelligente Sortierung...`
▸ Spanish	`Clasificación inteligente...`
▸ Italian	`Ordinamento intelligente...`
▸ Russian	`Умная сортировка...`

When a text is selected, the `Smart Sort...` command also appears in the contextual menu (RIGHT-CLICK.)

7. Quick test

After installing the script, you are ready to sort any word list in the blink of an eye. Here is an example:

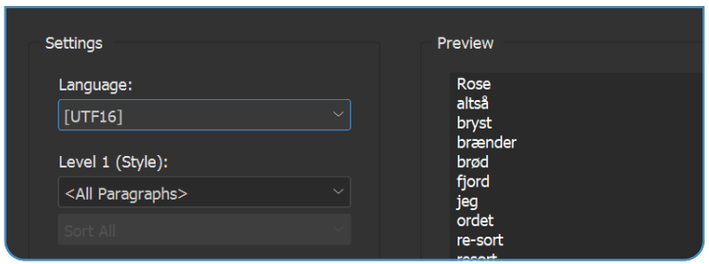
1) Open or create an InDesign document.

2) Copy the (unsorted) list below.

3) Create an empty text frame using the Type tool (T) and paste the lines from the clipboard.

4) Select the text frame and go into Type ▸ *Smart Sort*.

5) Browse the Language list and choose `[UTF16]` at the top. The Preview panel is updated so you can visualize how the words would be ordered according to the (basic) UTF-16 mapping.



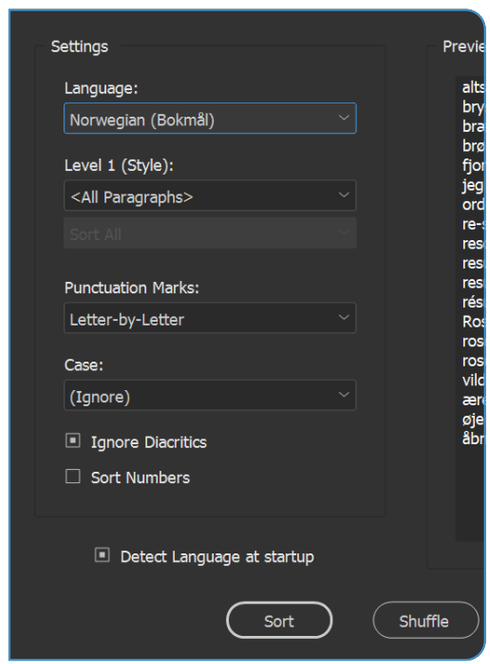
NOTE UTF-16 is a character encoding that essentially reflects how Unicode characters are ordered in the low ranges—in particular ASCII characters and Latin letters.

UNSORTED	[UTF16]	[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) Select `European Ordering Rules [EOR]` instead of `[UTF16]`. Note that the preview list is now ordered in a much more consistent way with respect to Latin digrams and accented letters (å, æ, é, ø.)

7) Finally, select the `Norwegian (Bokmål)` language and press the Sort button.

NOTE In the Dano-Norwegian alphabet, æ, ø and å are distinct letters and *collate* after z.



Language Settings

In text processing, the task of SORTING (or alphabetizing) relies on conventions that depends of course on the writing system, but a given ALPHABET does not entirely determine how words or expressions have to be ordered in a particular language, dialect or culture. Additional rules are needed to refine the process of *comparing strings*, referred to as COLLATING.

1. Supported scripts and languages

Basically, SmartSort recognizes letters (including accents, diacritics, ligatures) and punctuation marks used in about 230 languages. The script discriminates almost 10,000 characters and sequences extracted from the Default Unicode Collation Element Table (DUCET.) The below tables only show a subset of the Unicode blocks involved in the collation algorithm.

By design, SmartSort can deal with almost every known language based on Latin, Greek and/or Cyrillic alphabets. Regarding other writing systems like Hebrew, Arabic, Devanagari, Thai, Armenian, the script provides a minimal support inherited from the DUCET. However, the many languages and dialects that rely on those alphabets are not specifically "tailored."

NOTE As for unified ideographs and CJK-related characters, they are not supported in SmartSort.

TAILORING consists of applying custom changes to the default character ordering or to the collation algorithm. This mechanism is used to "provide linguistically-accurate collation," as discussed in the Unicode Collation Algorithm. A given writing system (say the Roman script) may have different rules when used for different languages. For example, in traditional Spanish the digram `ch` is considered a single letter and collates between `c` and `d` (the same happens with the digram `ll`.) Similarly, in Breton both the digram `ch` and the trigram `c'h` are considered own letters

UNICODE BLOCK	RANGE
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 essentially deals with Latin, Greek, and Cyrillic alphabets. Some additional writing systems are partially supported (Armenian, Hebrew, Arabic…)

UNICODE BLOCK	RANGE
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

Many other characters are investigated to handle diacritics, modifiers, punctuation, letterlike and alphanumeric symbols…

that separately collate between c and d. Conversely, ligatures like Æ, IJ, Œ, ß, are ordered respectively as AE, IJ, OE, ss in most European languages. (Those rules are managed from the generic [EOR] ordering.)

2. Multi-level comparison

SmartSort relies on a multi-level comparison mechanism inherited from the UCA.

▸ *Level 1 (base letter)* *E.g.* role < rule

In comparing two words, the main requirement is to order base letters like A vs. B whatever their particular case or additional marks, considered irrelevant at that level. This is known as LEVEL 1 COMPARISON and determined by the alphabet order of the target language. Note, however, that a base letter is not necessarily a single Unicode character (see the CH digram in Breton, the DŽ digram in Bosnian and Croatian, etc.) Also, what is seen as an accented character in some language may form a *distinct* base letter in another language: Ñ collates after N in Spanish, Å collates after Z in Swedish, etc.

▸ *Level 2 (accents and diacritics)* *E.g.* role << rôle

When two base letters are equivalent (like O and Ô in most Latin languages), the ordering can be fine-tuned based on accents or diacritical marks. This is known as LEVEL 2 COMPARISON. Some languages or cultures expect subtle adjustments at this level. For example, in German phonebooks the character ä is treated as a

LANGUAGE		TAILORING RULES *(overview)*
[EOR] European Ordering Rules	LATN/GREK	æ=ae ; œ=oe…
Afrikaans		n << ŉ
Albanian		c < ç ; d < dh … x < xh ; z < zh
Armenian		բ < և <<< ԵՒ
Asturian		*as **Spanish [Traditional]***
Azerbaijani	LATN/CYRL	c < ç ; e < ə … и < ы
Basque		*as **Spanish [Modern]***
Belarusian		e < ё ; y < ў
Bosnian	LATN/CYRL	c < č < ć ; d < dž < đ ; l < lj…
Breton		c < ch < c'h
Catalan		ll << l·l <<< L·l <<< L·L
Cebuano		*as **Filipino***
Croatian	LATN/CYRL	*as **Bosnian***
Czech		c < č ; h < ch…
Danish		d << đ … z < æ < ø < å…
Dutch		ij=ij
Dutch [Phonebook]		ij=y
English		æ=ae ; œ=oe
Esperanto		c < ĉ ; g < ĝ ; h < ĥ…
Estonian		s < š < z < ž ; v < õ < ä < ö…
Ewe		d < dz < ɖ ; e < ɛ ; f < ƒ…
Faroese		a < á ; d < ð … z < æ < ø…
Filipino		n < ñ < ng
Finnish		v << w ; y << ü … z < æ < ø…
Frisian [West]		i << y
Friulian		a < à < á < â ; e < è ; i < ì…
Galician		*as **Spanish [Traditional]***
German [Phonebook]		ae << ä ; oe << ö…
German [Phonebook/Austria]		a < ä ; o < ö ; u < ü…
Greenlandic		q << ĸ ; z < æ < ø < å…
Hausa *(boko)*	LATN *TRANSLIT. ARAB*	b < ɓ ; d < ɗ ; k < ƙ ; s < sh…
Hawaiian		a < e <<< E < i … w < ʻ
Hebrew		'~<GERESH> ; "~<GERSHAYIM>
Hungarian		c < cs ; d < dz < dzs ; g < gy…
Icelandic		a < á ; d < ð … z < þ < æ…
Igbo		b < ch ; g < gb < gh < gw…
Kannada		೦ < ೧ < ೨ < ೩ < ೪ < ೫ < ೬…
Kazakh		e < ё ; ұ < ү ; ы < і
Kurdish		c < ç ; e < ê ; i < î…

LANGUAGE		TAILORING RULES *(overview)*
Kyrgyz		e < ё
Latvian		c < č ; g < ģ … z < ž
Lingala		e < ɛ ; o < ɔ
Lingala [Phonetic]		e < ɛ ; o < ɔ ; g < gb ; k < kp…
Lithuanian		a << ą ; a < č … s < š…
Lower Sorbian		c < č < ć … k < ł…
Macedonian		ѓ < ѕ … ћ < ќ
Maltese		b < ċ ; f < ġ ; g < għ … y < ż
Moldavian	LATN/CYRL	*as **Romanian***
Northern Sami		a < á ; c < č … z < ž < æ…
Norwegian Bokmål		*as **Danish***
Oromo		z < ch < dh < kh < ny < ph < sh
Polish		a < ą ; c < ć … z < ź < ż
Romanian	LATN/CYRL	a < ă < â ; i < î ; s < ş ; t < ţ
Serbian	LATN/CYRL	*as **Bosnian***
Slovak		a < ä ; c < č ; h < ch … z < ž
Slovenian		c < č ; s < š ; z < ž
Spanish [Modern]		n < ñ
Spanish [Traditional]		c < ch ; l < ll ; n < ñ
Swedish		z < å < ä < ö << ø
Tatar		я < ә < ө < ү < җ < һ < ь
Tongan		n < ng ; z < ʻ …
Turkish		c < ç ; g < ğ ; h < ı ; o < ö…
Turkmen	LATN/CYRL	c < ç ; e < ä … e < ё ; ə < ә
Ukrainian		г < ґ ; я < ь…
Upper Sorbian		c < č ; d < dź ; e < ě…
Uzbek	LATN/CYRL	z < oʻ < gʻ < sh … e < ё
Vietnamese [Standard]		a < ă < â ; d < đ ; e < ê…
Welsh		c < ch ; d < dd…
Wolof		a < à ; e < é < ë ; n < ñ < ŋ…
Yoruba		e < ẹ ; g < gb ; o < ọ ; s < ṣ
Yoruba [Benin]		e < ɛ ; g < gb ; k < kp ; o < ɔ…

Special tailored languages supported in SmartSort 1.13 and quick overview of the associated collation rules.

Latin script · Cyrillic script · Mixed alphabets · Other writing system

variant of the digram ae but collates after it at level 2. (This is expressed by the rule ae << ä.) The same happens in Catalan with the forms ll and l·l.

▸ *Level 3 (case variants)* *E.g.* `bill <<< Bill`
Case differences (uppercase vs. lowercase) are typically ignored, *unless base letters and accents are identical.* A LEVEL 3 COMPARISON is then optionally processed depending on the particular conventions adopted in your document.

NOTE Level 3 rules only discriminate words that are already equivalent at level 1 and 2. For example, the option "Lowercase first" will produce `ball <<< Ball < bill <<< Bill < bull`, while "Uppercase first" leads to `Ball <<< ball < Bill <<< bill < bull`.

3. Choosing your language

In order to get your headings sorted with respect to the appropriate rules, the first step is to choose a language.

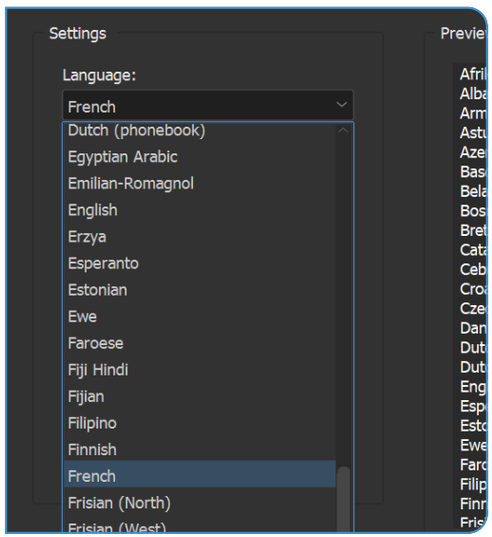
1) Select the target paragraphs (or table rows.)

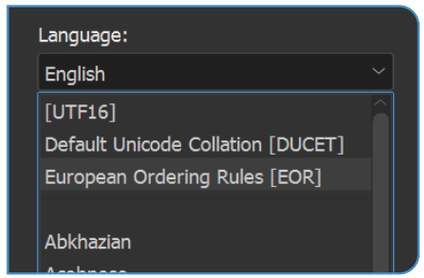
2) Invoke SmartSort from the Type menu (Type ▸ *Smart Sort…*)

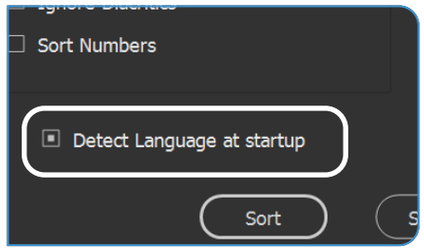
3) Click the Language dropdown list and select the desired option. More than 230 items are available here, including special systems like *Dutch (phonebook)*, *German (phonebook)*, *Norwegian Bokmål* vs. *Norwegian Nynorsk*, etc. If no language fits your needs, select either

- `European Ordering Rules [EOR]` if the writing system is limited to Latin/Cyrillic/Greek scripts;
- `Default Unicode Collation [DUCET]` if foreign alphabets are used and should be treated as specified in the DUCET;
- `[UTF16]` if no specific language is involved at all (like in product codes, ASCII names…)

NOTE If the option "Detect Language at startup" is turned on, the script automatically suggests a language based on the selection.





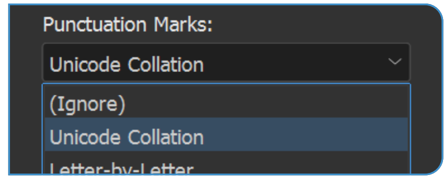


4. Punctuation marks

Spaces, hyphens, commas and hundreds of other punctuation marks are known as *variable collation elements*. Unlike extra characters that are purely ignored by the algorithm, variable elements may impact ordering. The Unicode collation table assigns special weights to variable elements so they can be sorted consistently. The default ordering is summarized below:

1. Spaces.
2. Hyphens and dashes.
3. Comma, semicolon, colon.
4. Exclamation and question marks.
5. Full stop and dots.
6. Quotation marks.
7. Parentheses and brackets.
8. Other signs.

The Punctuation Marks dropdown list offers four options: *(Ignore)*, *Unicode Collation*, *Letter-by-Letter*, and *Word-by-Word*.



- *(Ignore)*
  Select this option to get variable elements ignored while sorting the list. The items are then alphabetized *as if no punctuation marks were present*. Hence, forms like `InDesign`, `In Design`, `in-design`, `(InDesign)`, `#indesign`… are all equivalent—insofar as case differences don't matter.

- *Unicode Collation*
  This option assigns weights to variable elements as specified in the DUCET. Rules like `[SPACE]` < `[HYPHEN]` < `[PARENTHESIS]` will then apply.

(Ignore)	Unicode Collation
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

Examples of either ignoring or applying Unicode rules to punctuation marks.
• The yellow blocks (left frame) indicate equivalent expressions.
• The highlighted characters (right frame) show elements that affect the order.

Symbol	Name
»	TAB
·	SPACE
^	NO-BREAK SPACE
	EN QUAD
	FLUSH SPACE
	EN SPACE
	EM SPACE
	THIRD SPACE
	QUARTER SPACE
	SIXTH SPACE
#	FIGURE SPACE
!	PUNCT. SPACE
	THIN SPACE
	HAIR SPACE
	MATH SPACE
‾	OVERLINE
_	LOW LINE
‗	DBLE LOW LINE
-	HYPHEN MINUS
‐	HYPHEN
‒	FIGURE DASH
–	EN DASH
—	EM DASH
―	HORIZ. BAR
, ، ‚ …	COMMA incl. Armenian and Arabic commas, decimal and thousands separators…
;	SEMICOLON
؛	ARAB. SEMICOLON
:	COLON
! ‼ ¡ ? ¿ …	EXCLAMATION AND QUESTION MARKS…
.	FULL STOP
․	ONE DOT LEADER
‥	TWO DOT LEADER
…	HORIZ. ELLIPSIS
·	MIDDLE DOT
‘ ’ ‚ ‛ ‹ › “ ” „ ‟ « »	QUOTATION MARKS
( ) [ ] { }	PARENTHESES AND BRACKETS
§	SECTION
¶	PILCROW
@	COMMERCIAL AT
*	ASTERISK
/	SOLIDUS
\	REVERSE SOLIDUS
&	AMPERSAND
#	NUMBER SIGN
%	PERCENT SIGN
‰	PER MILLE SIGN
†	DAGGER
‡	DOUBLE DAGGER
°	DEGREE SIGN
҂	CYRIL. THOUSANDS
+	PLUS
±	PLUS-MINUS
÷	DIVISION
×	MULTIPLICATION
<	LESS-THAN
=	EQUALS
>	GREATER-THAN
|	VERTICAL LINE
¦	BROKEN BAR
~	TILDE

Small subset of variable elements ordered by increasing weight. This collation order applies when Unicode Collation is selected.

▸ *Letter-by-Letter*

The LETTER-BY-LETTER system ignores variable elements but parentheses and commas used as separators. According to the *Chicago Manual of Style,* "The order of precedence is one word, word followed by a parenthesis, word followed by a comma, then (ignoring spaces and other punctuation) word followed by a number, and word followed by letters." [2010, p. 833.]

As in the *(Ignore)* option, alphabetization proceeds across spaces and skips most punctuation marks, with hyphenated expressions being treated as one word. The only difference is that parentheses and commas interrupt the lead term.

NOTE The rule [PARENTHESIS] < [COMMA] is implemented in accordance with the *Chicago Manual of Style*, although not universally accepted across publishers. For example, the *Oxford Guide to Style* gives the ordering `High, J.` < `high (light-headed)` as correct. Common in the USA, the letter-by-letter system is mostly used in encyclopedias, atlases, and dictionaries.

▸ *Word-by-Word*

The WORD-BY-WORD system ignores variable elements but parentheses, commas, spaces or hyphens used as separators. According to the *Oxford Guide to Style,* "The two parts of a hyphenated compound are treated as separate words, except where the first element is not a word in its own right." [2002, p. 581.] Hence hyphens should be ignored in special cases and treated as spaces otherwise. Since this cannot be decided from the collation algorithm, SmartSort considers any hyphen as equivalent to a space.

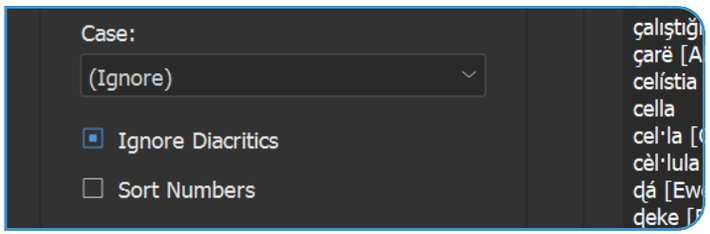
(Ignore)	Letter-by-Letter	Word-by-Word
New, Arthur	NEW (Neighbors Ever Watchful)	NEW (Neighbors Ever Watchful)
newborn	NEW (Now End War)	NEW (Now End War)
newcomer	New, Arthur	New, Arthur
New Deal	New, Zoe	New, Zoe
new economics	newborn	New Deal
newel	newcomer	new economics
New England	New Deal	New England
"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)

Spaces and hyphens then collate after parentheses and commas, so `A(B` < `A,B` < `A B`. In addition, spaces and hyphens are discarded if followed by a left parenthesis or a comma, so `A (B` is equivalent to `A(B` and still collates before `A,B`.

NOTE British publishers prefer the word-by-word system. Many alphabetization standards also recommend it, especially for indexes and "related information retrieval devices."

Letter-by-Letter collation ignores punctuation marks except (and , which operate as lead term breakers.

Word-by-Word collation treats spaces and hyphens as separators too. This tends to keep lead terms together.



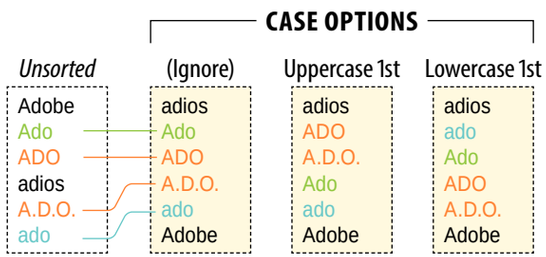
5. Case and diacritics

Options relative to CASE and DIACRITICS only focus on items *that are equivalent at level 1.* In many cases this won't impact the order at all, since most terms are already alphabetized due to level 1 differences (base letters.)

NOTE Keep in mind, however, that a letter modified by a diacritic may be treated as a distinct letter (rather than a diacritic variant) depending on the language you have selected. For example, the forms CARE and ÇARË are totally distinct in Albanian, where C < Ç.

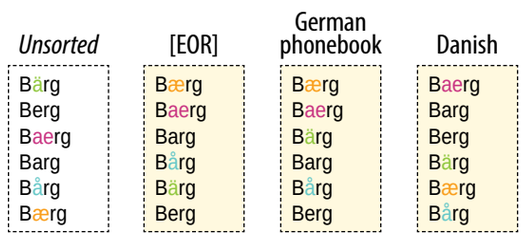
- *Case*
  – Select *Uppercase first* to make uppercase variants come first, as in `adios < A.D.O <<< Ado <<< ado < Adobe`.
  – Select *Lowercase first* to make lowercase variants come first, as in `ball < bill <<< Bill <<< BILL < bull`.
  – Select *(Ignore)* to disregard level 3 differences.

- *Diacritics*
  Check *Ignore Diacritics* to wash away level 2 differences. This will keep diacritical variants (like ÈRES and ÉRÈS) in their original order. In many dictionaries and indexes, however, rules on accents and diactrics should be applied with respect to cultural conventions. Here are a few ones that SmartSort implements:
  – The Europen Ordering Rules [EOR] treat `ə` (resp. `ɪ`, `ʞ`, `ŋ`, `ŧ`) as a level 2 variant of `e` (resp. `i`, `k`, `n`, `t`.) In addition, `þ` is regarded as a level 2 variant of `th` (`th << þ`.)
  – In Catalan, `ll << l·l` (which anyway is the default [EOR] ordering, but this contrasts with Traditional Spanish.)
  – In Finnish, `w` is a foreign letter that tends to be treated as a level 2 variant of `v` (`v << w`.)
  – In German phonebooks, `ä` (resp. `ö` and `ü`) is treated as a level 2 variant of `ae` (resp. `oe` and `ue`.)



Case sorting is a subtle preference that decides how equivalent terms (e.g. ADO ≈ ado) must be internally arranged. All case variants keep their relative position in the *(Ignore)* table, because SmartSort performs a *stable sort.* When *Uppercase first* is selected, the forms ADO and A.D.O.* collate before Ado, which itself collates before ado. This order is reversed when *Lowercase first* is choosen.

* ADO and A.D.O. remain equivalent unless Unicode "Punctuation Marks" are in use.



LEVEL2 COLLATION deals with accents, diacritics and various related forms: ae vs. ä in German phonebooks, ä vs. æ in Danish, etc.

6. Digits and numbers

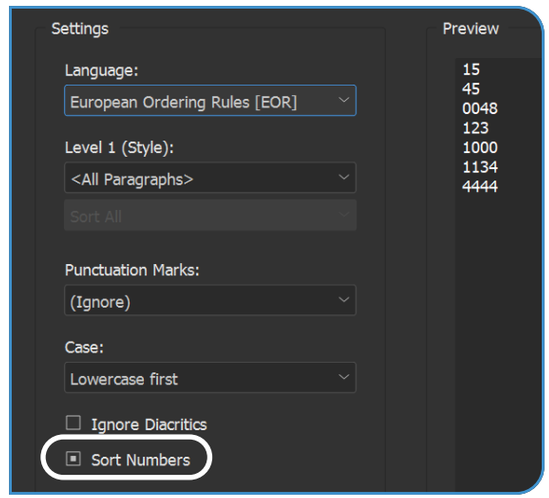
By default, digit characters are ordered as single collation elements with respect to their associated weight. They collate *after* punctuation marks and *before* letters. Equivalent Unicode digits like 1, ¹ (U+00B9), ١ (U+0661), १ (U+0967), ₁ (U+2081), etc, have the same primary weight. SmartSort handles such characters as specified in the DUCET—unless you select [UTF16].

The *Sort Numbers* option specifically targets INDEPENDENT DIGIT SEQUENCES formed of characters among 0123456789. Foreign digits are ignored, and only *separate* terms are parsed. For example, the expressions *"2001, a Space Odyssey"* and "InDesign CC 2022" contain valid sequences (2001, 2022), while "ABC123" provides no value—a word boundary is required to start a valid sequence.

NOTE As well as other language options, *Sort Numbers* is disabled if [UTF16] is selected. Characters are then sorted per code point, which may still make sense in some cases. However, the resulting order is clearly desappointing regarding numeric values:

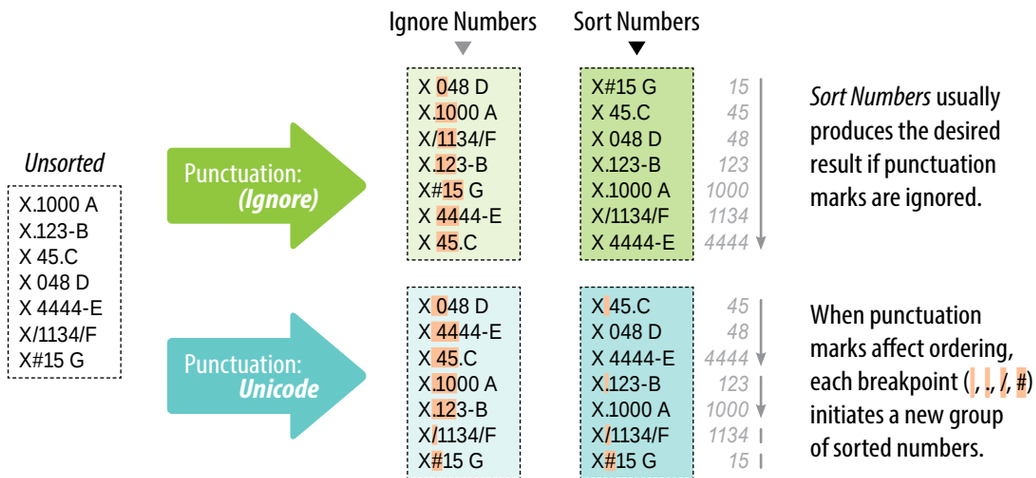
Unsorted	[UTF16]
1000	0048
123	1000
45	1134
0048	123
4444	15
1134	4444
15	45

Unsorted	Sort Numbers
1000	15
123	45
45	0048
0048	123
4444	1000
1134	1134
15	4444



▸ *Sorting simple numbers*
If your list only contains digits, choose any language, e.g. [EOR], and check *Sort Numbers*. The digit sequences are then intepreted as numbers and sorted accordingly.

▸ *Sorting complex expressions*
If the digit sequences are mixed with other characters, the effect of *Sort Numbers* depends on whether punctuation marks interact with ordering. Different settings are illustrated below.



Sort Numbers usually produces the desired result if punctuation marks are ignored.

When punctuation marks affect ordering, each breakpoint (, ., /, #) initiates a new group of sorted numbers.

Text Processing

SmartSort has been designed to sort a set of paragraphs as fast as possible, preserving inner styles and other text attributes. The current version supports multiple text frames and table rows. It can also perform a TWO-LEVEL SORT based on paragraph styles.

1. Preview area

SmartSort's dialog provides a Preview area on the right side so the user can visualize the effect of changing language options before s/he clicks the **Sort** button.

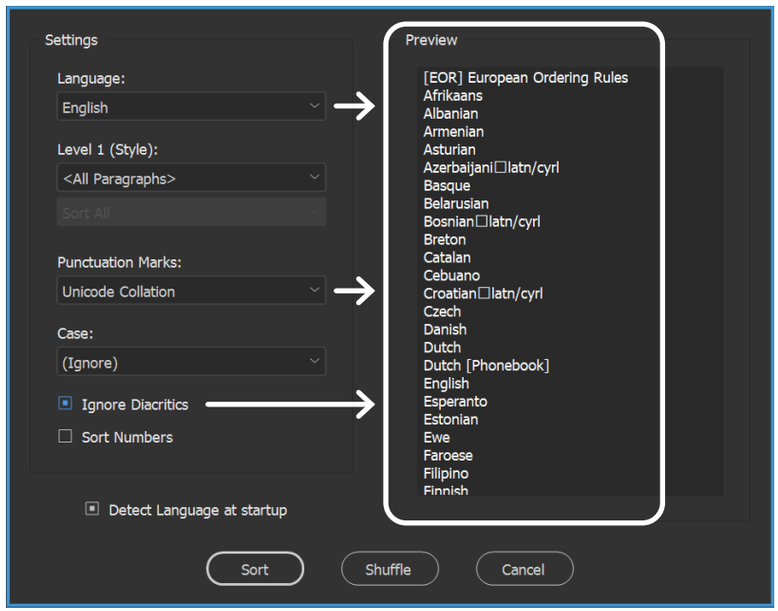
If your InDesign selection contains many paragraphs, only a sublist of about 100 lines is loaded in memory and only the top elements (from 20 to 25 lines) are visible in the Preview area. Alphabetization is applied to the sample list and updated whenever an option is changed.

2. Selecting the target paragraphs

The SELECTED PARAGRAPHS determine the incoming data, each paragraph being considered a distinct item.

1) Open your InDesign document.

2) Go to the text container (frame, cell) that hosts your list.



The Preview area displays the top items of the list once ordered according to the current options. Special characters that cannot be rendered here (e.g RIGHT INDENT TAB) are indicated by a square.

3) Using the Type Tool (T), select the range of paragraphs to be reordered. At least two paragraphs are required, meaning that the text selection must contain at least one ¶.

NOTE If a paragraph is partially selected, it is anyway loaded and processed entirely. *(See the figure below.)*

NOTE Forced line breaks (¬) are not recognized as item separators. If your list is formatted this way, use Find/Replace to change ^n into ^p.

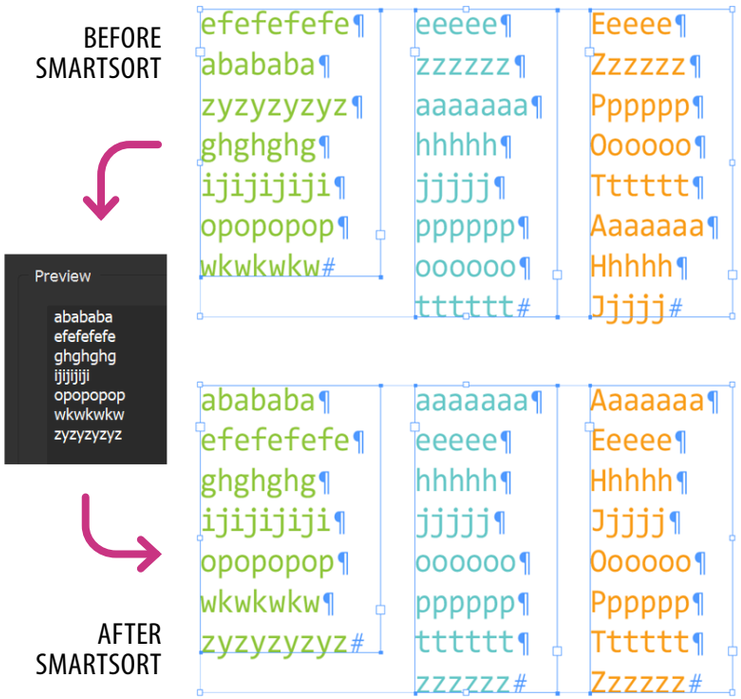
4) Invoke SmartSort going into Type ▸ *Smart Sort…*—or using the contextual menu RIGHT-CLICK ▸ *Smart Sort…*

Selection (InDesign)
Target paragraphs (SmartSort)

3. Targeting full text containers

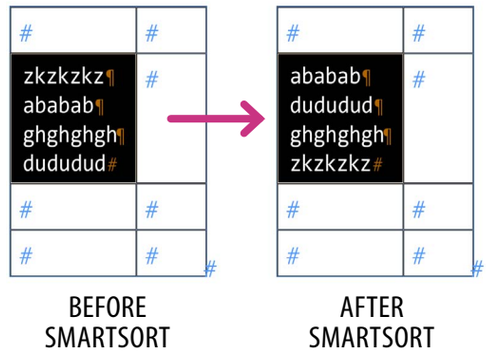
- *Single text frame* → *story*
  If you select a text frame (Selection Tool), the entire story is considered, including overset text.

- *Multiple text frames* → *multiple stories*
  If you select two or more text frames, all associated stories are considered (incl. overset text) and processed *separately*. In this case the Preview area only shows items extracted from the first-selected frame.



- *Table cell(s)* → *inner paragraphs*
  – If you select a cell with the Type Tool, the inner paragraphs are considered (incl. overset text.)
  – If multiple cells are selected, each is considered a distinct story with its own list, so the corresponding sets of paragraphs are processed separately. This policy *only applies if the average number of paragraphs per cell is greater than one.*

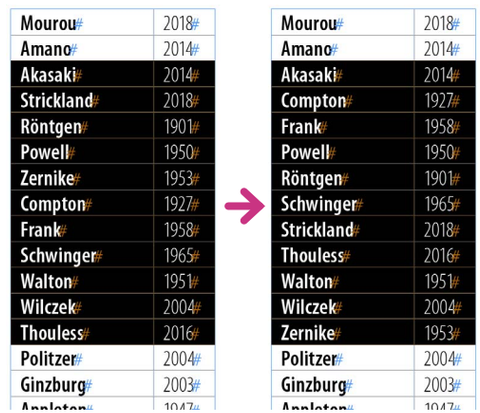
NOTE In all cases examined here, you can invoke SmartSort from Type > *Smart Sort. . .*, or from the contextual menu.



4. Sorting table rows

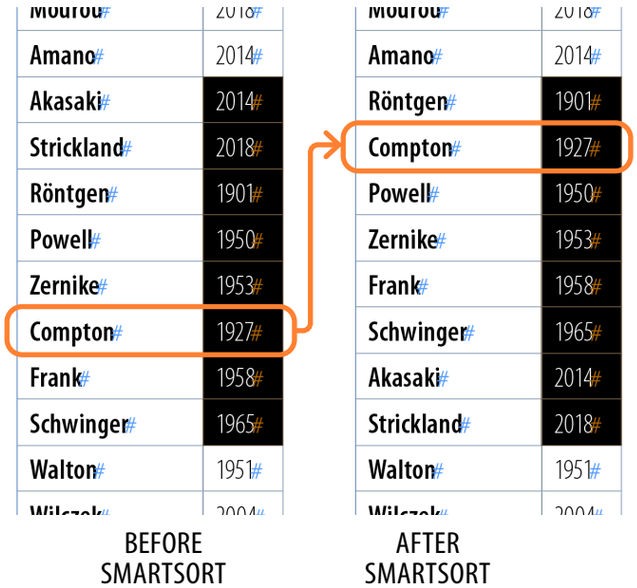
A range of TABLE ROWS can be sorted as would be a set of paragraphs, each row being seen as a line. ***This special feature is highly time-consuming.*** It only supports simple table structures (no splitted or merged cells), provided that each target cell contains *one paragraph on average.*

Select the rows that need to be ordered and run the script. The lead terms are then extracted from the first column. If



The selected rows are sorted based on names found in the first column.

you need to sort rows *with respect to another column*, select only the cell range that contains the SORT KEYS:



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#

BEFORE SMARTSORT AFTER SMARTSORT

Note that the target rows are still processed as a whole. For example, the row "`Compton | 1927`" moves to the 2nd place.

NOTE Processing rows from an InDesign script is much slower than processing paragraphs. If hundreds of cells are involved, it may take several tens of seconds to complete the task.

NOTE If you select an entire table all body rows are considered.

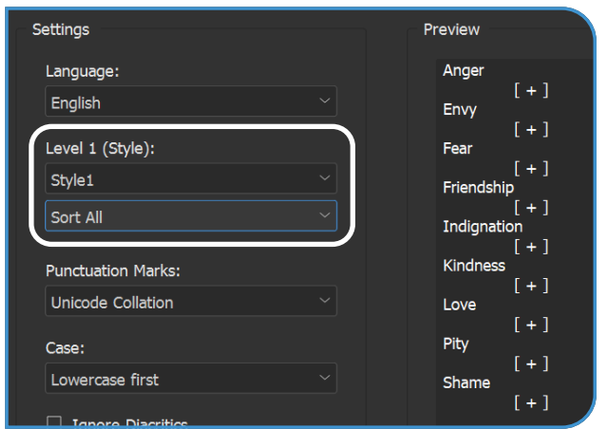
5. Two-level sort

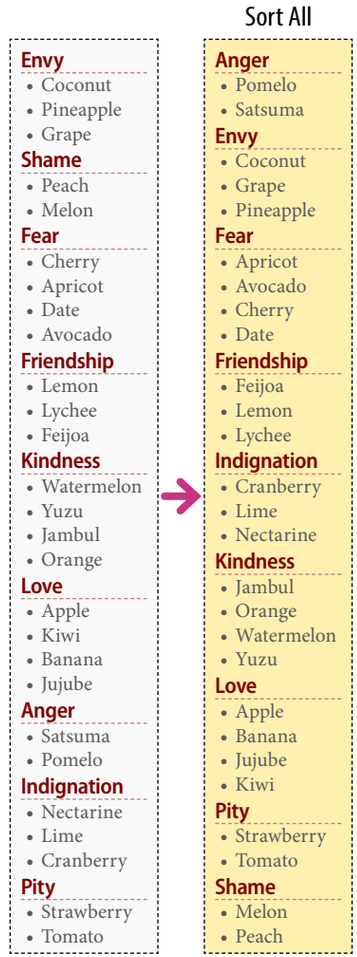
Given a structured document based on headings, titles, etc, with associated PARAGRAPH STYLES, it may be relevant to apply sorting at a particular LEVEL while keeping the text flow consistent at other levels.

If the target text involves *at least two paragraph styles*, SmartSort allows you to declare one particular style as the *Level 1 (Style)* and then offers three distinct options:

▸ *Sort All*
Sorts both level 1 headings against each other, and each group of paragraphs under that level.

NOTE The Preview area only displays the head terms (sorted); the symbol [+] indicates sub-items being sorted too.

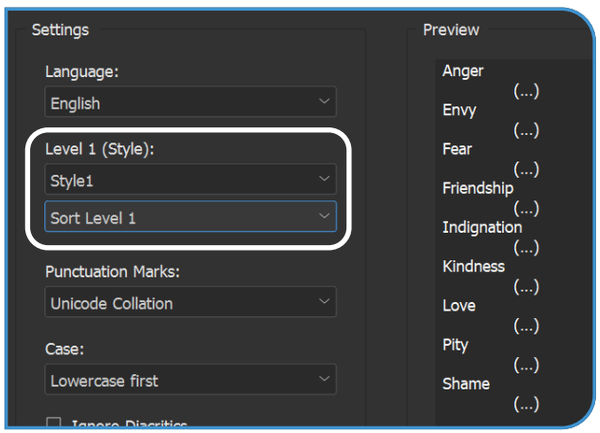




▶ *Sort Level 1*

Sorts headings against each other, keeping the underlying groups of paragraphs unchanged.

NOTE The Preview area then only displays the head terms (sorted) and the symbol **(...)** indicates sub-items being unchanged.



This option is useful to alphabetize the entries of a dictionary, each head term having a dedicated style applied, followed by one or many paragraph(s).

Sort Level 1

Envy
- Coconut
- Pineapple
- Grape

Shame
- Peach
- Melon

Fear
- Cherry
- Apricot
- Date
- Avocado

Friendship
- Lemon
- Lychee
- Feijoa

Kindness
- Watermelon
- Yuzu
- Jambul
- Orange

Love
- Apple
- Kiwi
- Banana
- Jujube

Anger
- Satsuma
- Pomelo

Indignation
- Nectarine
- Lime
- Cranberry

Pity
- Strawberry
- Tomato

→

Anger
- Satsuma
- Pomelo

Envy
- Coconut
- Pineapple
- Grape

Fear
- Cherry
- Apricot
- Date
- Avocado

Friendship
- Lemon
- Lychee
- Feijoa

Indignation
- Nectarine
- Lime
- Cranberry

Kindness
- Watermelon
- Yuzu
- Jambul
- Orange

Love
- Apple
- Kiwi
- Banana
- Jujube

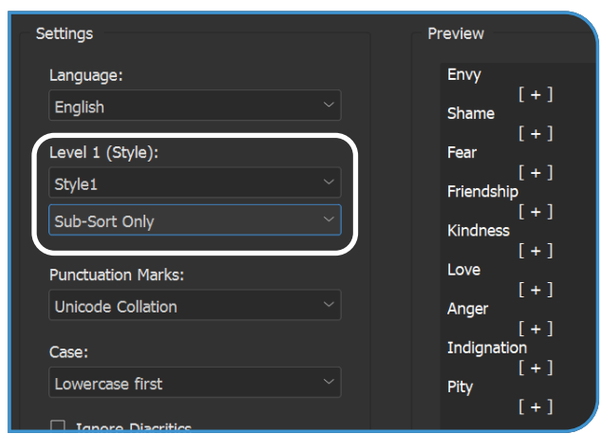
Pity
- Strawberry
- Tomato

Shame
- Peach
- Melon

▶ *Sub-Sort Only*

Keeps the headings in their original order and sorts the underlying paragraphs.

NOTE The Preview area then only displays the head terms (unsorted) and the symbol **[+]** indicates sub-items being sorted.



NOTE If a *flat* sort is needed (instead of one of the mechanisms detailed above), simply select `<All Paragraphs>` in the *Level 1 (Style)* list.

Sub-Sort Only

Envy
- Coconut
- Grape
- Pineapple

Shame
- Melon
- Peach

Fear
- Apricot
- Avocado
- Cherry
- Date

Friendship
- Feijoa
- Lemon
- Lychee

Kindness
- Jambul
- Orange
- Watermelon
- Yuzu

Love
- Apple
- Banana
- Jujube
- Kiwi

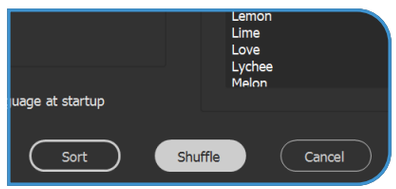
Anger
- Pomelo
- Satsuma

Indignation
- Cranberry
- Lime
- Nectarine

Pity
- Strawberry
- Tomato

6. Shuffling

Instead of sorting a list (**Sort** button), you can redistribute the items randomly by clicking **Shuffle**. This function obviously ignores all language settings.



Last Words

SmartSort is a freeware and will remain so. It took many years to implement the underlying algorithm in the spirit of the Unicode specification. The present version is light enough to operate in the InDesign scripting engine. It should help you solve common—and even uncommon—sorting issues.

As it has been primarily designed for addressing a flat list (of paragraphs), SmartSort cannot handle *every* possible text structure, like complex tables or multilevel indexes. Also, despite the many languages and settings it supports, there are still a lot of cases where additional refinements should be done, with respect to specific conventions, cultures, etc.

The processing of simple tables has been added recently. Due to technical limitations of the scripting subsystem, this feature often leads to high execution time. This is a typical case where a better, custom implementation may exist and fit your particular needs.

SmartSort is a relatively unknown tool although it provides an original solution to a common problem, namely, *sorting styled data in InDesign* (with respect to language rules.) If you have any interest in using this script, please let us know and do not hesitate to report bugs.

Also, if you want to support the development of other valuable InDesign scripts, give a look at our products at indiscripts.com.

SmartSort 1.1

www.indiscripts.com

A plug-in for Adobe® InDesign® based on Adobe® ExtendScript and ScriptUI. Created, designed and developed by Marc Autret. User Interface available in English, French, German, Spanish, Italian, and Russian.

My special thanks to Peter Kahrel and Jean-Claude Tremblay, who greatly helped me improve, question and test SmartSort under various conditions.

Main Page: http://indiscripts.com/category/projects/SmartSort
Download: http://indiscripts.com/blog/public/scripts/SmartSort.zip
Bugs and Feedback: support@indiscripts.com

End User License Agreement: http://indiscripts.com/pages/eula
Copyright Notice: http://indiscripts.com/pages/copyright