

Justified Arabic text with stretchable kashida in LaTeX

This post describes how to make stretchable pseudo-*kashidas* and how to automatically insert these at appropriate letter connections in order to justify Arabic texts, that is, to give it even right and left margins. The problem, solution, and the result is first presented in a non-technical way. Hereafter the implementation in LaTeX is then discussed in some detail.

Introduction

A paragraph with even left and right margins is said to be *justified*. Of course. In texts in the Latin alphabet, this is achieved in two ways: a) by hyphenating long words where they appear at the end of a line, and b) by varying the space between words.¹ In the Arabic script there is no hyphenation, words cannot be broken up in two, making the justification of paragraphs more difficult.

What Arabic has, though, that the Latin script lacks, is the ability stretch out the connecting lines between letters. In typed text, this is done with a character called *kashida* (U+0640). It is a line at the baseline, similar in shape to the underscore, and it connects to adjacent letter. The word *kabīr* ‘big’, for instance, could be written with a few kashidas to make it take up more space, () which can be used justify the text.

There are two problems with this. Firstly, any line can only be lengthened by a multiple of the width of the kashida in the typeface. For example, if the kashida is 5 pts wide, the line could be lengthened by 5 pts by adding a kashida somewhere, or by 10 pts by adding two. But it could not be lengthened by 7 pts. Secondly, the lengthening can only be applied point wise certain words on the line and cannot be distributed evenly across the line, making for an uneven appearance of the words in line.

Se for example the excerpt below from the novel *Yūtūbiyā* by Aḥmad Xālīd Tawfīq.[[^]utopia]. Kashidas have been added to the last couple of words at

¹There is also a third method; to stretch or squeeze letters by a measure that is too small to be noticeable on individual letters, but with an accumulated effect that provides a variability used for justifying the text. See the documentation for the `mycrotypes` package for a discussion and implementation in LaTeX.

the end of most lines to fill them up to the left margin. This makes for an unattractive and uneven page.

لا تقاوم يا أحمق! ... ما الذي تمنحك إياه لحظات أخرى من
العيش مع الأغيار؟ .. ما الذي لم تحققه في سنواتك العشرين
السابقة وتتوي أن تحققه لو ظللت حيًا؟ ... فرارك هذا لا يختلف
عن فرار الصرصور على جدار مطبخ، أو أميبا تنزلق تحت
عدسة مجهر ... صرخة غريزة لا أكثر .. إنه تفاعل التحاشي
الذي زرعه الطبيعة فيك، وعليك أن تتعلم كيف تهمله كي تظفر
براحة استحققتها...
انطلقت الرشاشات فنظر لأعلى .. نعم .. هذه الطلقات من
أجلك أنت ..
ترسم ذلك الخط الطويل على الرمال .. الخط الذي يمر بك
أنت ...

Figure 1: Page 12 from *Yūtūbiyā*, where kashidas are used in words at the end of the line to fill it.

Solution - the stretchable kashida

One solution to this is to use a stretchable kashida that is placed between all connecting characters. It just so happens that TeX provides to do this fairly easily and reliably. If no stretching is needed to justify the text it gets a width of zero meaning that it does not appear at all. If some filling is needed to fill the line (which in practice will virtually always be the case), then all these stretchable kashidas are stretched with an equal amount to fill the line. There is thus no fixed or even default width of this kashida, it will be as long or short as it needs be, and all stretchable kashidas on the same line will be of the same width.

Have a look at the page below. It contains the same three paragraphs, taken at random from an Arabic news site, typeset three times, first in three columns, then in four, and then in five columns. The more columns, and thus the narrower the columns, the more difficult it is for the typesetting engine to find a good way of distributing words on the lines to avoid big white spaces between words. In the last part with five columns, the text looks really bad, with large white blobs scattered all over surface of the text.

Now compare this with the following page. Here the same text is set with the same font, font size and columns, but with a stretchable kashida at every letter

شنت القوات العراقية هجوماً على مطار مدينة الموصل، أحد أهم أهداف العملية العسكرية لطرد مسلحي تنظيم الدولة الإسلامية من شطر المدينة الغربي. كما اقتحمت القوات معسكر الغزلاني القريب من المطار، والذي شهد اشتباكات عسكرية عنيفة، بحسب وكالة أسوشيتد برس. يشار إلى أن تنظيم الدولة قد دمر بالفعل ممر الطائرات، غير أن الاستيلاء على مثل هذا الموقع المهم من شأنه مساعدة الجيش العراقي في السيطرة على الطرق المؤدية إلى المدينة.

استعادت القوات العراقية والميليشيات الموالية لها الشطر الشرقي للمدينة، الشهر الماضي. ويقول مراسل بي بي سي، كوينتين سومرفيل، الذي رافق القوات العراقية، إنهم وصلوا إلى محيط المطار. وقال متحدث باسم الجيش لفضائية "العراقية" إن غارات متزامنة يُجرى شنها على المطار وقاعدة الغزلاني لـ "تشتيت" مسلحي تنظيم الدولة. وكانت منشورات قد أُلقيت في وقت سابق، للتحذير من شن هجوم وشيك منازلتهم داخل المدينة وحولها.

على غرب المدينة، الذي يقول مسؤولون عسكريون إن شوارع المتعرجة والضيقة قد تجعل استعادة المنطقة مهمة صعبة. وعلى الرغم من أن غرب المدينة أصغر بقليل من شرقها، إلا أنه أكثر كثافة ويضم مناطق يُنظر إليها باعتبارها مؤيدة لتنظيم الدولة. وأعربت الأمم المتحدة عن قلقها بشأن الأوضاع المعيشية للمدنيين المحاصرين في المدينة، وسط تقارير عن أن عددهم قد يرتفع إلى 650 ألفاً. وفرّ بالفعل ما يربو على 160 ألف شخص من منازلهم داخل المدينة وحولها.

شنت القوات العراقية هجوماً على مطار مدينة الموصل، أحد أهم أهداف العملية العسكرية لطرد مسلحي تنظيم الدولة الإسلامية من شطر المدينة الغربي. كما اقتحمت القوات معسكر الغزلاني القريب من المطار، والذي شهد اشتباكات عسكرية عنيفة، بحسب وكالة أسوشيتد برس. يشار إلى أن تنظيم الدولة قد دمر بالفعل ممر الطائرات، غير أن الاستيلاء على مثل هذا الموقع المهم من شأنه مساعدة الجيش العراقي في السيطرة على الطرق المؤدية إلى المدينة.

استعادت القوات العراقية والميليشيات الموالية لها الشطر الشرقي للمدينة، الشهر الماضي. ويقول مراسل بي بي سي، كوينتين سومرفيل، الذي رافق القوات العراقية، إنهم وصلوا إلى محيط المطار. وقال متحدث باسم الجيش لفضائية "العراقية" إن غارات متزامنة يُجرى شنها على المطار وقاعدة الغزلاني لـ "تشتيت" مسلحي تنظيم الدولة. وكانت منشورات قد أُلقيت في وقت سابق، للتحذير من شن هجوم وشيك منازلتهم داخل المدينة وحولها.

على غرب المدينة، الذي يقول مسؤولون عسكريون إن شوارع المتعرجة والضيقة قد تجعل استعادة المنطقة مهمة صعبة. وعلى الرغم من أن غرب المدينة أصغر بقليل من شرقها، إلا أنه أكثر كثافة ويضم مناطق يُنظر إليها باعتبارها مؤيدة لتنظيم الدولة. وأعربت الأمم المتحدة عن قلقها بشأن الأوضاع المعيشية للمدنيين المحاصرين في المدينة، وسط تقارير عن أن عددهم قد يرتفع إلى 650 ألفاً. وفرّ بالفعل ما يربو على 160 ألف شخص من منازلهم داخل المدينة وحولها.

شنت القوات العراقية هجوماً على مطار مدينة الموصل، أحد أهم أهداف العملية العسكرية لطرد مسلحي تنظيم الدولة الإسلامية من شطر المدينة الغربي. كما اقتحمت القوات معسكر الغزلاني القريب من المطار، والذي شهد اشتباكات عسكرية عنيفة، بحسب وكالة أسوشيتد برس. يشار إلى أن تنظيم الدولة قد دمر بالفعل ممر الطائرات، غير أن الاستيلاء على مثل هذا الموقع المهم من شأنه مساعدة الجيش العراقي في السيطرة على الطرق المؤدية إلى المدينة.

استعادت القوات العراقية والميليشيات الموالية لها الشطر الشرقي للمدينة، الشهر الماضي. ويقول مراسل بي بي سي، كوينتين سومرفيل، الذي رافق القوات العراقية، إنهم وصلوا إلى محيط المطار. وقال متحدث باسم الجيش لفضائية "العراقية" إن غارات متزامنة يُجرى شنها على المطار وقاعدة الغزلاني لـ "تشتيت" مسلحي تنظيم الدولة. وكانت منشورات قد أُلقيت في وقت سابق، للتحذير من شن هجوم وشيك منازلتهم داخل المدينة وحولها.

على غرب المدينة، الذي يقول مسؤولون عسكريون إن شوارع المتعرجة والضيقة قد تجعل استعادة المنطقة مهمة صعبة. وعلى الرغم من أن غرب المدينة أصغر بقليل من شرقها، إلا أنه أكثر كثافة ويضم مناطق يُنظر إليها باعتبارها مؤيدة لتنظيم الدولة. وأعربت الأمم المتحدة عن قلقها بشأن الأوضاع المعيشية للمدنيين المحاصرين في المدينة، وسط تقارير عن أن عددهم قد يرتفع إلى 650 ألفاً. وفرّ بالفعل ما يربو على 160 ألف شخص من منازلهم داخل المدينة وحولها.

Figure 2: Example text without stretchable kashida.

connection. Note the total lack of white blobs. The stretchable kashida nicely fills every line, producing an even surface.

NOTE: When viewed on screen, the inserted stretchable kashida may appear choppy in certain resolutions. This is because it has not been adapted for screen display in various resolutions, as has the rest of the font. Zoom in on the pdf-image and you will see that the choppiness disappears. It will also not be an issue when printed.

Application

The rest of this post is step wise an explanation of the LaTeX code that achieves this result. A complete self-contained example (used to produce the documents above) can be found in this [GitHub repo](#).

First we load the packages that we need and setup the Arabic font. Lateef is a free and open source font available [here](#).

```
\documentclass{article}
\usepackage{calc} % used to measure kashida thickness
\usepackage{polyglossia}
\setmainlanguage{arabic}
\newfontfamily\arabicfont[Script=Arabic]{Lateef}
```

XeTeXinterchartoks

Then we activate the `XeTeXinterchartoks` feature, declare the different classes, and assign characters to the different classes. This allows us to automatically insert a command between combinations of characters of each class. Each class is assigned with characters with a specific set of connecting rules, i.e. those that connect to the following letter and those that do not. So between all connectors we want to insert the command for the stretchable kashida, also when a connector is followed by a non-connector, but not after a non-connector. We also need special classes for the letters *lām* () and *alif* ,) in all its variants) since we do not want to insert anything between them when they appear in this order, which would break the compulsory ligature . Similar excepting for other ligatures in particular fonts could easily be added this way, otherwise the ligatures will simply connect at the baseline as normal when this code is applied.

```
\XeTeXinterchartokenstate=1

\newXeTeXintercharclass\confb % connect back
\newXeTeXintercharclass\conb % connect front back
\newXeTeXintercharclass\alif % alif
\newXeTeXintercharclass\lam % lam
```

شنت القوات العراقية هجوماً على مطار مدينة الموصل، أحد أهم أهداف العملية العسكرية لطرد مسلحي تنظيم الدولة الإسلامية من شطر المدينة الغربي. كما اقتحمت القوات معسكر الغزلاني القريب من المطار، والذي شهد اشتباكات عسكرية عنيفة، بحسب وكالة أسوشيتد برس. يشار إلى أن تنظيم الدولة قد دمر بالفعل ممر الطائرات، غير أن الاستيلاء على مثل هذا الموقع المهم من شأنه مساعدة الجيش العراقي في السيطرة على الطرق المؤدية إلى المدينة.

واستعادت القوات العراقية والميليشيات الموالية لها الشطر الشرقي للمدينة، الشهر الماضي. ويقول مراسل بي بي سي، كوينتين سومرفيل، الذي رافق القوات العراقية، إنهم وصلوا إلى محيط المطار. وقال متحدث باسم الجيش لفضائية "العراقية" إن غارات متزامنة يُجرى شنها على المطار وقاعدة الغزلاني لـ "نشتيت" مسلحي تنظيم الدولة. وكانت منشورات قد أُلقيت في وقت سابق، للتحذير من شن هجوم وشيك على

غرب المدينة، الذي يقول مسؤولون عسكريون إن شوارعها المتعرجة والضيقة قد تجعل استعادة المنطقة مهمة صعبة. وعلى الرغم من أن غرب المدينة أصغر بقليل من شرقها، إلا أنه أكثر كثافة ويضم مناطق يُنظر إليها باعتبارها مؤيدة لتنظيم الدولة. وأُعلنت الأمم المتحدة عن قلقها بشأن الأوضاع المعيشية للمدنيين المحاصرين في المدينة، وسط تقارير عن أن عددهم قد يرتفع إلى 650 ألفاً. وفّر بالفعل ما يربو على 160 ألف شخص من منازلهم داخل المدينة وحولها.

شنت القوات العراقية هجوماً على مطار مدينة الموصل، أحد أهم أهداف العملية العسكرية لطرد مسلحي تنظيم الدولة الإسلامية من شطر المدينة الغربي. كما اقتحمت القوات معسكر الغزلاني القريب من المطار، والذي شهد اشتباكات عسكرية عنيفة، بحسب وكالة أسوشيتد برس. يشار إلى أن تنظيم الدولة قد دمر بالفعل ممر الطائرات، غير أن الاستيلاء على مثل هذا الموقع المهم من شأنه مساعدة الجيش العراقي في السيطرة على الطرق المؤدية إلى المدينة.

واستعادت القوات العراقية والميليشيات الموالية لها الشطر الشرقي للمدينة، الشهر الماضي. ويقول مراسل بي بي سي، كوينتين سومرفيل، الذي رافق القوات العراقية، إنهم وصلوا إلى محيط المطار. وقال متحدث باسم الجيش لفضائية "العراقية" إن غارات متزامنة يُجرى شنها على المطار وقاعدة الغزلاني لـ "نشتيت" مسلحي تنظيم الدولة. وكانت منشورات قد أُلقيت في وقت سابق، للتحذير من شن هجوم وشيك على غرب المدينة، الذي يقول مسؤولون عسكريون إن شوارعها المتعرجة والضيقة قد تجعل استعادة المنطقة مهمة صعبة. وعلى الرغم من أن غرب المدينة أصغر بقليل من شرقها، إلا أنه أكثر كثافة ويضم مناطق يُنظر إليها باعتبارها مؤيدة لتنظيم الدولة. وأُعلنت الأمم المتحدة عن قلقها بشأن الأوضاع المعيشية للمدنيين المحاصرين في المدينة، وسط تقارير عن أن عددهم قد يرتفع إلى 650 ألفاً. وفّر بالفعل ما يربو على 160 ألف شخص من منازلهم داخل المدينة وحولها.

شنت القوات العراقية هجوماً على مطار مدينة الموصل، أحد أهم أهداف العملية العسكرية لطرد مسلحي تنظيم الدولة الإسلامية من شطر المدينة الغربي. كما اقتحمت القوات معسكر الغزلاني القريب من المطار، والذي شهد اشتباكات عسكرية عنيفة، بحسب وكالة أسوشيتد برس. يشار إلى أن تنظيم الدولة قد دمر بالفعل ممر الطائرات، غير أن الاستيلاء على مثل هذا الموقع المهم من شأنه مساعدة الجيش العراقي في السيطرة على الطرق المؤدية إلى المدينة.

واستعادت القوات العراقية والميليشيات الموالية لها الشطر الشرقي للمدينة، الشهر الماضي. ويقول مراسل بي بي سي، كوينتين سومرفيل، الذي رافق القوات العراقية، إنهم وصلوا إلى محيط المطار. وقال متحدث باسم الجيش لفضائية "العراقية" إن غارات متزامنة يُجرى شنها على المطار وقاعدة الغزلاني لـ "نشتيت" مسلحي تنظيم الدولة. وكانت منشورات قد أُلقيت في وقت سابق، للتحذير من شن هجوم وشيك على غرب المدينة، الذي يقول مسؤولون عسكريون إن شوارعها المتعرجة والضيقة قد تجعل استعادة المنطقة مهمة صعبة. وعلى الرغم من أن غرب المدينة أصغر بقليل من شرقها، إلا أنه أكثر كثافة ويضم مناطق يُنظر إليها باعتبارها مؤيدة لتنظيم الدولة. وأُعلنت الأمم المتحدة عن قلقها بشأن الأوضاع المعيشية للمدنيين المحاصرين في المدينة، وسط تقارير عن أن عددهم قد يرتفع إلى 650 ألفاً. وفّر بالفعل ما يربو على 160 ألف شخص من منازلهم داخل المدينة وحولها.

Figure 3: Example text with stretchable kashida.

[illegible]

```
\XeTeXcharclass ~ \=\alif
\xeTeXcharclass ~ \=\alif
\XeTeXcharclass ~ \=\alif
\xeTeXcharclass ~ \=\alif
```

Then we declare the combinations of character classes between which we want to insert the command for the stretchable kashida (which we have yet to define). The order is significant. Note that all possible combination of characters classes where you want stretchable kashida must be added to the list.

```

\XeTeXinterchartoks \lam \confb = {\kashida{}}
\XeTeXinterchartoks \lam \conb  = {\kashida{}}
\XeTeXinterchartoks \confb \conb = {\kashida{}}

```

This means that the insertion of a vowel sign (, etc), *shadda* (), *sukūn* (), or other similar symbol negates the stretchable kashida, since the sequence is the ****[vowel sign]—letter,*** and not one of the sequences listed above. You can see this in the word , the first word in the test paragraph in the examples above.

Kashida thickness

The following snippet uses functions from the `calc` package to measure the thickness of the kashida in the loaded font. It measures the top of the kashida from the baseline and stores this value in `\kashidaheight`. It also measures how far it extends below the baseline and stores the value in `\kashidadepth`. (None of the fonts I have tested this on have a kashida that extends below the baseline, so this measure will normally be 0 pts.)

```

\newlength\kashidaheight
\setlength\kashidaheight{\heightof{\textarabic{}}}
\newlength\kashidadepth
\setlength\kashidadepth{\depthof{\textarabic{}}}

```

This means that this method does not give good results for typefaces in which the letter connection is raised above the baseline. This is the case Geeza Pro typeface that is included in OSX. (This could be fixed provided there is a reliable way of measuring how much the letter connections, and by extension the kashida character of the font, is raised above the baseline.) The method presented here will work fine with most simple *nasx* typefaces, however, such as Lateef, Scheherazadw, and Simplified Arabic (distributed with MS Word), as well as with any mono-spaced Arabic font.

Note also that these measurements will change any time the font size changes, so that the above command should be run again after any command that changes the fontsize. Alternatively, and preferably, these commands could be redefined to run `\setlength\kashidaheight{\heightof{\textarabic{}}}` and `\setlength\kashidadepth{\depthof{\textarabic{}}}`.

Defining the kashida command

With the measurements of the kashida now stored, we can use them to define the actual stretchable kashida that is to go between all connecting characters.

```

\newcommand\kashida[1]{\char"200D
  \nobreak\leaders
    \hrule height \kashidaheight depth \kashidadepth

```

```
\hskip 0pt plus 100 pt
\nobreak\char"200D}
```

This produces a leader, similar to the row of dots that fills the line between a chapter title and the page number in a table of contents, for example, but it is here in the form of a solid line of kashida-thickness.

Note that the command begins and ends with `\char"200D`. This inserts the Unicode character `ZERO WIDTH JOINER`, a character that is very useful when typing in Arabic. As its name suggest, it takes no space and only instructs surrounding letters to connect with it, and thus with oneanother. Without it the typesetting engine would see the command `\nobreak` when it comes to this command, and not a letter with which the previous letter can connect, making it take the unconnecting form. The command `\nobreak` instructs LaTeX not to have any line breaks here.

The heart of the command is `\leaders` which is a stretchable element, here declared in plain TeX. It takes two parameters:

- a) what it is that is to be repeated or stretched, here a horizontal rule (`\hrule`) with the height stored in `\kashidaheight` and the depth stored in `\kashidadepth`
- b) the stretchability (or ‘glue’ in TeX-terminology), here set to be able to stretch from 0 to 100 pts.

It is important that the glue is set to limited dimension such as 100 pts and not to infinite glue with `fill`, because this will mess up the last line of the paragraph that does not have to be filled.

Conclusion

And that’s it. Adding the above code to the preamble automatically applies the stretchable kashida, as in the multicolumn examples above.

The way it has been applied here comes with a few caveats:

- It only works with typefaces where letters are connected at the baseline and where the connection is not cruvilinear.
- If the typeface have ligatures other than *lam-alif* that you wish to retain, they must be added to the classes and patters of `XeTeXinterchartoks`.
- Vowel signs and similar negate the stretchable kashida.
- Your may not be able to resist the urge to hang your beautiful Arabic text columns on your office wall.