

Arabic text justification

Mohamed Jamal Eddine BENATIA,
Mohamed ELYAAKOUBI and Azzeddine LAZREK
Department of Computer Science,
Faculty of Science, University Cadi Ayyad
P.O. Box 2390, Marrakesh, MOROCCO
Phone: +212 24 43 46 49 Fax: +212 24 43 74 09
lazrek (at) ucam dot ac dot ma
<http://www.ucam.ac.ma/fssm/rydarab>

Abstract

Justification of Latin script based texts is carried out by handling of hyphenation and insertion of inter-word glue, which can be stretched or shrunk to some extent. Due to the cursive nature of Arabic writing, text justification in Arabic has a quite different logic. So, the classical algorithms of text justification must be completely revised. Justification in Arabic typography has traditional processes inspired by calligraphy manuals. Words are flexible: on the one hand, a word can be resized through stretching some letters in a curvilinear way; on the other hand, it can be shrunk via ligatures so that the width of a given letter group is decreased.

محاذاة النص العربي

ملخص: تعتمد برامج معالجة النصوص عند اختيار محاذاة نص لاتيني تقطيع الكلمات المجاورة للهامش وإدخال فراغات قابلة للتمطيط والتقليص بشكل متناسب بين الكلمات.

في الكتابة العربية الممشوقة، تتصل الحروف ببعضها، الشيء الذي يجعل لمحاذاة النص العربي أسلوبا مختلفا تماما. ويحتم مراجعة لوغاريتمات المحاذاة بالنسبة للتبولوجيا العربية. تستند محاذاة النصوص في الكتابة العربية إلى أساليب مستقاة من كتب الخط العربي، بحيث تمتاز الكلمات بمرونتها، فهي من جانب قابلة للمط من خلال تمديد الحروف بشكل انسيابي بواسطة الكشيدة. ومن جانب آخر، قابلة للضغط من خلال تراكب الحروف و تداخلها عموديا، الشيء الذي يقلص العرض الذي تحتله الكلمة على السطر.

1 Justification in Latin typography

A good many methods to make paragraphs on a page be visually homogeneous have been developed. The majority of these methods have already been implemented in \TeX . The hyphenation and justification algorithm divides a paragraph into lines in an optimal way, as regards time complexity as well as the visual result obtained [6]. The processing spreads out beyond the paragraphs, to reach the level of the page in its totality.

1.1 Typographical hyphenation

Typographical hyphenation is the breaking of words when they come at the end of a line and would overflow into the margin. In general, word breaking happens at syllable boundaries.

In the beginning, the hyphenation of words was done by hand. In 1983, F. M. Liang [8] published a sophisticated method to find nearly all the suitable

places to insert hyphens in a Latin script based written word. The method is controlled by an organized tree structure of *tries*, containing a list of hyphenation patterns. Combinations of letters which allow, or prohibit, the word-breaking are listed, and priorities to breakpoints in letter groups are assigned. Patterns reflect the hyphenation rules of a given language. So there will be as many pattern tree structures as there are normative languages. This is the algorithm Donald E. Knuth chose to implement in \TeX .

1.2 Spacing

To give a line the flexibility it needs, a space or its equivalent can presumably be inserted between each pair of words. Therefore, some of these spaces are transformed into line ends. Others are transformed into variable sized spaces called glue. The glue has a normal size that can be stretched or shrunk. When a paragraph intended to have a justified right margin

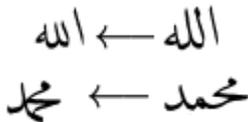


Figure 1: Special morphology



Figure 2: An Arabic word with kashida

is composed with $\text{T}_{\text{E}}\text{X}$, the glue widths of each line are adjusted so that lines end almost at the right margin. Generally, the last line of a paragraph is an exception, and does not have to end at the margin.

Given this, it is always an extremely difficult task to obtain a uniform typographical gray. The main reason is the impossibility of ensuring an equal inter-word space in different lines. The composition accidents of *rivers* and *alleys* are uncomfortable for the reader, and irregular spaces catch his attention.

A solution for such problems has been implemented by Hàn Th  Thành [11]. Instead of (or as well as) changing inter-word spaces to justify text lines, the widths of characters are slightly modified. So, better inter-word spacing can be obtained and space elasticity can be limited. This width modification is implemented through horizontal scaling of fonts in pdf $\text{T}_{\text{E}}\text{X}$. If it is employed parsimoniously and wisely, this method can appreciably improve appearance of the typography produced by $\text{T}_{\text{E}}\text{X}$.

2 Justification in Arabic typography

Arabic writing is cursive in its printed form as well as in its handwritten form. The letters' morphology changes according to their position in the word, according to the surrounding letters, and in some cases, according to the word's meaning (for example, ALLAH and Mohamed when it indicates the prophet's name in Figure 1). The alternative positions then depend on the typeset words. The end of a given glyph is tied to the beginning of the following glyph, with no possible break.

Remark:

All figures and the table are written from right to left according to Arabic writing direction.

2.1 Kashida

The genuine connections between Arabic letters are curvilinear strokes that can be stretched or shrunk according to the writing context. Such curve is called *kashida*, *tamdid*, *madda*, *maT*, *taTwil*, or *iTalah*. This variable-sized connection between letters is specific to Arabic alphabet based writing. The kashida is used in various circumstances:

- **emphasis**: to mark an important piece of a word. The kashida will then mark the sound elongation;
- **legibility**: to give a better character layout on the baseline, and to lessen the cluttering at the joint point between two successive letters of the same word;
- **aesthetics**, to embellish the writing of a word;
- **justification**, to justify a text line.

The example in Figure 2 shows a composition of an Arabic word; the arrows indicate the kashidas, with various degrees of extensibility.

There are mandatory elongations, allowed elongations and prohibited elongations. The typographical quality of a text is determined, among other things, by the absence of mandatory elongations or the presence of prohibited elongations.

In terms of Arabic text justification, the kashida is a typographical effect that allows the lengthening of letters in some carefully selected points of the line, with determined parameters, in order to produce the left alignment of a paragraph. The good selection of kashidas' insertion points is called *tansil*.

2.2 Current typesetting systems

In terms of text processing tools, the curvilinear kashida is, generally, still beyond what the majority of typesetting systems can afford. The kashida is not a character in itself, but an elongation of some character parts while keeping rigid the body's character. It is not a simple horizontal scaling to widen character width. Instead of performing a kashida, the majority of typesetting systems proceed by inserting a rectilinear segments between letters; the resulting typographical quality is unpleasant. Due to the lack of adequate tools, the solution consists of inserting a glyph, that is, an element of a font. So, rather than computing (say) parameterized B zier curves in real time, a ready-to-use character is inserted. Moreover, whenever stretching is performed by means of a parameterized glyph coming from an external dynamic font, the current font context is changed.

Curvilinear extensibility of characters can be afforded by certain systems through the a priori generation of curvilinear glyphs for some predefined sizes.

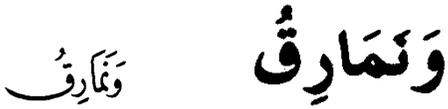


Figure 3: Presence of Noon-Meem ligature

(Source: Holy Quran written by the calligrapher Alhaj Hafez Mohamed Amin Alrochdi, scrutinized and revised by General Directorate of Endowments, Baghdad, p. 649.)



Figure 4: Absence of Noon-Meem ligature

Beyond these sizes, the system will choose curvilinear primitive and linear fragments. Of course, this will violate the curvilinear shape of letters and symbols composed at large sizes [3, 9, 10].

A better approach consists of building a dynamic font [4, 7], through parameterizations of the composition procedure of each letter. The introduced parameters indicate the extensibility size or degree. To handle the elongations, a letter is decomposed into two parts: the static body of the letter and the dynamic part, capable of stretching.

2.3 Ligatures

The cursive nature of Arabic writing implies, among other things, a wide use of ligatures [5]. Indeed, Arabic writing is rich in ligatures. Some ligatures are mandatory and obey grammatical and contextual rules [5]. Others are optional and exist only for aesthetic reasons, legibility and/or justification. Moreover, the connection of letters, in the course of writing cursively, can lead to the introduction of implicit contextual ligatures. An explicit ligature is the fusion of two, three, or even more, graphemes.

Some aesthetic ligatures result in some reading ambiguity. So, precise texts such as the Holy Quran are sometimes written without such ligatures. Figure 3 shows an example of such ambiguity: the Noon-Meem ligature can be confused with the initial form of the letter Ghain. Figure 4 shows the text without the ligature.

Generally, the ligature width is shorter than the width of the fused grapheme group. For example, the aesthetic ligature نم in Figure 5 is 9.65031pt wide, whereas the ligature given by simple contextual substitutions نح is rather wide (14.754pt).

The control of ligature behavior, by the conversion of implicit ligatures into aesthetic ones, brings some flexibility to the word. So, it can be adapted to the available space on the line. The example in Fig-



Figure 5: Contextual and aesthetic transformations

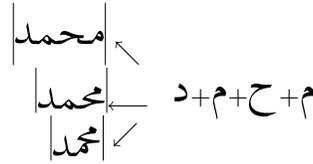


Figure 6: Various levels of ligatures

ure 6 shows three ligature levels: mandatory simple substitutions, aesthetic ligatures of second degree, and finally, aesthetic ligatures of third degree. The two last ligature levels provide shrinking possibilities of the same word.

The use of aesthetic ligatures of second and third degree has to take into consideration the constraints of legibility.

A typesetting system should take into account three levels of recourse to ligatures. In the first level, there are only implicit contextual ligatures and mandatory grammatical ligatures of second degree. This level is recommended for textbook composition, where it is necessary to avoid any collision between characters and/or any reading ambiguity. In a second level, some aesthetic ligatures of second degree can be used. This level is recommended for composition of books for the general public. The third level, where the use of aesthetic ligatures of higher degrees is allowed and liberties in graphic expressions can be taken, is possible in special circumstances.

The use (or not) of the explicit ligatures to improve the justification should take into account the graphic environment and the block regularity of the concerned text. In calligraphy, when an aesthetic ligature is used, there is no obligation to use this ligature in all text occurrences. The justification problem can be resolved by the use of kashidas in a text containing only implicit ligatures.

The use of ligatures to justify lines is not solely an Arabic writing characteristic. Adolf Wild [12], the Gutenberg museum conservator in Mainz, examined the Gutenberg Bible from a typographical point of view. At the level of lines, justification was made via ligatures, instead of today's variable spaces.

2.4 Other practices

The baseline (*satr alkitabah*) is the line on which letters stand to form words and therefore phrases. In Arabic writing, the baseline is a virtual line. The

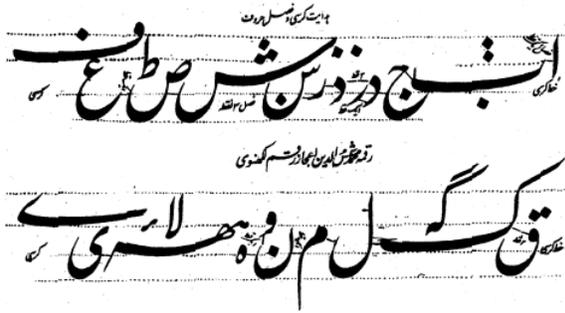


Figure 7: Kursive in ta'lyq style (Source: Badai' Alkhat Al'arabi, p. 372, Naji Zine Eddine Almasref)

characters as well as the words do not stay directly on a static line with a fixed reference. The letters flow in an interactive way with the surrounding letters. They change not only their shape, but also their position with respect to an imaginary baseline. Letters as well as words are free from any absolute reference. Thereby, the letters as well as the words may be tangled. According to Ibn Muqlah [1], good composition implies three types of interacting forces among letters and words:

- *tarsif*, the junction of the conjoint letters, where each letter seeks the point of join with the preceding letter;
- *ta'lif*, the addition of letters in isolated form, where the isolated letter seeks to fill the blank left by the typeface of their neighboring letters;
- *tastir*, the words' ruling, so that they are held horizontally.

From another standpoint, positioning variations of letters and words are at the origin of the *chairs kursy-s* [2] theory, that introduces a number of bearing lines. According to different versions of this theory, the number of these lines can vary from two to five, or even seven lines. The heights between lines vary according to authors of this theory and can not be respected in practice. Apparently, this theory is used more in style ta'lyq, as in Figure 7. Thereby, we can wonder to what extent we can formalize this concept for a possible automation thereof.

Nevertheless, we can say that for justification, calligraphers do use the tangling possibilities of letters and words, as in Figure 8, and inter-word space shortening as additional or alternative methods instead of kashidas. The same text was written by two calligraphers and won a prize in the IRCICA¹ 2004 competition. The composition in Figure 9 by

¹ <http://www.ircica.org/>

قَالَ الْإِمَامُ عَلِيُّ بْنُ أَبِي طَالِبٍ كَرَّمَ اللَّهُ وَجْهَهُ فِي وَصْفِ الدُّنْيَا

Figure 8: Words tangle

قَالَ الْإِمَامُ عَلِيُّ بْنُ أَبِي طَالِبٍ كَرَّمَ اللَّهُ وَجْهَهُ فِي وَصْفِ الدُّنْيَا
 أَمَا بَعْدُ فَأَيُّ حَذْرٍ كَرَّمَ اللَّهُ الدُّنْيَا فَأَنَّهَا حُلُوهٌ خَصِيْرَةٌ حُمِفَتْ بِالشَّهَوَاتِ وَتَحَبَّتْ
 بِالْعَاجِلَةِ وَرَاقَتْ بِالْعَلِيلِ وَتَحَلَّتْ بِالْأَمَالِ وَتَزَيَّنَتْ بِالْعُرُورِ لِأَنَّ دَوْمَ
 حَزَنُهَا وَلَا تُمْرُ مِنْ جَعْفَتِهَا عَزْرَةٌ صَرَّارَةٌ حَاشِلَةٌ زَائِلَةٌ نَافِذَةٌ بَاطِلَةٌ
 أَسْكَلَةُ عَوَالَةٍ لَا تَعْدُو. إِذَا تَنَاهَتْ إِلَى أُنْبِيَاءِ أَهْلِ الرَّغْبَةِ فِيهَا وَالرِّضَاءِ بِهَا
 أَنْزَلَتْكُمْ رُكُومًا قَالَ اللَّهُ تَعَالَى سُبْحَانَهُ. كَمَا أَنْزَلْنَا مِنَ السَّمَاءِ مَا نَخَلَطُ بِهِ
 نَبَاتَ الْأَرْضِ فَأَصْبَحَ هَشِيمًا تَذْرُوهُ الرِّيحُ وَكَانَ اللَّهُ عَلَى كُلِّ شَيْءٍ مُقْتَدِرًا
 لَسَيْفِكُمْ أَنْ تَمُرُّ مِنْهَا فِي حَبْرَةٍ إِلَّا أَعْقَبْتُمْ بِمَدَاهَا عِبْرَةٌ وَلَا يَلِيقُ فِي سِرَائِهَا
 بَطْنًا إِلَّا أَعْقَبْتُمْ مِنْ صَرَائِهَا ظَهْرٌ وَلَسَتْ تَطْلُقُ فِيهَا دِيمَةٌ رَخَاءٌ إِلَّا أَعْقَبْتُمْ عَلَيْهِ
 مُرْتَبَةٌ بَلَاءٌ وَحَرِيْرٌ إِذَا أَصْبَحَتْ لَهُ مُنْصَرَّةٌ أَنْ تُشْفِيَ لَهُ مُشْكِرَةٌ وَإِنْ حَاطَتْ بِهَا

Figure 9: Justification favoring the use of increasing spaces

قَالَ الْإِمَامُ عَلِيُّ بْنُ أَبِي طَالِبٍ كَرَّمَ اللَّهُ وَجْهَهُ فِي وَصْفِ الدُّنْيَا
 أَمَا بَعْدُ فَأَيُّ حَذْرٍ كَرَّمَ اللَّهُ الدُّنْيَا فَأَنَّهَا حُلُوهٌ خَصِيْرَةٌ حُمِفَتْ بِالشَّهَوَاتِ وَتَحَبَّتْ بِالْعَاجِلَةِ
 وَرَاقَتْ بِالْعَلِيلِ وَتَحَلَّتْ بِالْأَمَالِ وَتَزَيَّنَتْ بِالْعُرُورِ لِأَنَّ دَوْمَ حَزَنُهَا وَلَا تُمْرُ مِنْ جَعْفَتِهَا عَزْرَةٌ
 صَرَّارَةٌ حَاشِلَةٌ زَائِلَةٌ نَافِذَةٌ بَاطِلَةٌ أَسْكَلَةُ عَوَالَةٍ لَا تَعْدُو. إِذَا تَنَاهَتْ إِلَى أُنْبِيَاءِ أَهْلِ الرَّغْبَةِ فِيهَا
 وَالرِّضَاءِ بِهَا أَنْزَلَتْكُمْ رُكُومًا قَالَ اللَّهُ تَعَالَى سُبْحَانَهُ. كَمَا أَنْزَلْنَا مِنَ السَّمَاءِ مَا نَخَلَطُ بِهِ
 نَبَاتَ الْأَرْضِ فَأَصْبَحَ هَشِيمًا تَذْرُوهُ الرِّيحُ وَكَانَ اللَّهُ عَلَى كُلِّ شَيْءٍ مُقْتَدِرًا لَسَيْفِكُمْ أَنْ تَمُرُّ مِنْهَا
 فِي حَبْرَةٍ إِلَّا أَعْقَبْتُمْ بِمَدَاهَا عِبْرَةٌ وَلَا يَلِيقُ فِي سِرَائِهَا بَطْنًا إِلَّا أَعْقَبْتُمْ مِنْ صَرَائِهَا ظَهْرٌ
 وَلَسَتْ تَطْلُقُ فِيهَا دِيمَةٌ رَخَاءٌ إِلَّا أَعْقَبْتُمْ عَلَيْهِ مُرْتَبَةٌ بَلَاءٌ وَحَرِيْرٌ إِذَا أَصْبَحَتْ لَهُ مُنْصَرَّةٌ أَنْ تُشْفِيَ لَهُ
 مُشْكِرَةٌ وَإِنْ حَاطَتْ بِهَا
 أَيْدِي دَوْبٍ وَأَجَلُو لِي أَنْ تَمُرُّ مِنْهَا عِبْرَةٌ رَغْبًا إِلَّا أَرَقْتُمْ مِنْ فُرَائِهَا
 قَبِيْرًا وَلَا يُمْسِي مِنْهَا فِي جَنَاحِ أَمْرٍ إِلَّا أَمْرٌ عَلَى قَوَادِمِ حَرْفٍ حَرَارَةٌ تُعْرُومُ فِيهَا فَابِيَةٌ فَإِنْ مَرَّ عَلَيْهَا
 لِأَخِيْرٍ فِي شَيْءٍ يَوْمًا رَوْدًا هَذَا إِلَّا أَلْفَعُو مِنْ أَمَلٍ مِنْهَا أَسْتَسْكِرُ فَرِيْرًا يَوْمِيْنَهُ وَمَنْ أَسْتَسْكِرُ فَرِيْرًا

Figure 10: Justification favoring the use of kashida

M. T. Alubaidy used more spacing and entanglement, while A. Alabdo in Figure 10 favored more kashida.

Calligraphers also build on other practices for justification, such as:

- word heaping; this consists of putting certain words above others — especially, the word Allah above the preceding word (see the end of the seventh line of Figure 11);



Figure 11: Heaping words and hyphenating fragments

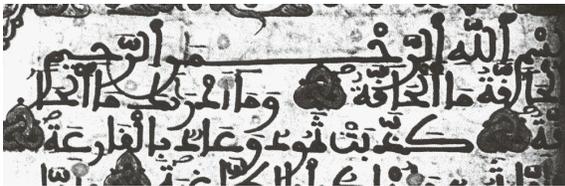


Figure 12: Word hyphenation

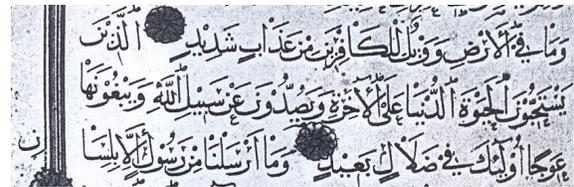


Figure 13: Expulsion in margin of hyphenated fragment (Source: Holy Quran written by Yakout Almusta'simy with style Ryhany and Kufy in Iran museum)

- moving the broken fragment above the hyphenated word, as in the last word of the fourth line in Figure 11;
- word hyphenation, as in Figure 12;
- word hyphenation in margin; this is an expulsion of hyphenated fragments to the line's margin, instead of the following line, as in Figure 13;
- decreasing of some words at the end of a line, as in Figures 14 and 15;
- curving of the baseline, as in Figure 16.

Since the 10th century, hyphenations at ends of lines have been strictly prohibited. This was probably due to the cursive structure of the Arabic language, and to the possible absence of vocalization signs. There are even sentences which it is advised not to hyphenate (ex. *salla ALLAH elyh w sallam*). If word hyphenation was accepted in Arabic, legibility would be considerably affected.

Nowadays, only the famous caesura at hem-stitch boundary is allowed. Regarded as a pause,

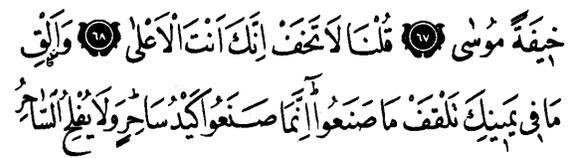


Figure 14: Decreasing of words at end of line



Figure 15: Decreasing of words at end of line (Source: Badai' Alkhat Al'arabi, p. 372, Naji Zine Eddine Almasref)

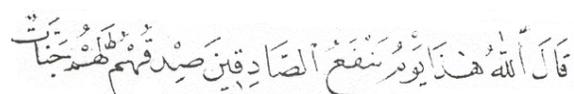


Figure 16: Curving of the baseline

أَيُّهَا السَّائِلِي عَنِ الصَّدِّاقِ وَالظَّامِ
 إِنَّ حِفْظَ الصَّلَاةِ يُغْنِيكَ فَاسْمَعْنَهَا اسْتِمَاعَ امْرِئٍ لَهُ اسْتِيقَاةٌ
 هِيَ غَمِيضَةٌ وَالْمَظَالِمُ وَالْأَنْظِلُ سَلَامٌ وَالظُّلْمُ وَالظُّبَى وَاللِّحَاةُ
 وَالْعَطَا وَالظَّكِيمُ وَالظُّبَى وَالنَّعِيَّةُ عَظْمٌ وَالظُّلُّ وَاللِّتْنَى وَالشَّوَابَةُ
 وَالنَّظْمِيُّ وَالنَّظْمُ وَالنَّظْمُ وَالنَّظْمُ رِيضَةٌ وَالْقَيْظُ وَالظُّكْمَا وَاللِّحَاةُ

Figure 17: Caesura at hemstitch boundary

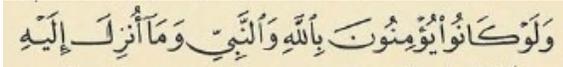


Figure 18: Stretching Lam

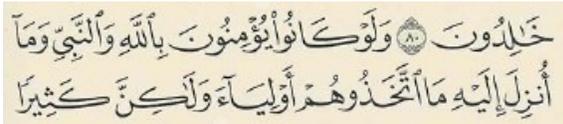


Figure 19: Simple Lam

this does not affect legibility since the hyphenated word is still held on the same line, as shown in Figure 17. There is also a stretching letter, shown in Figure 18, which is not allowed anymore, Figure 19.

2.5 Diacritic/metric dot

The qalam used to produce the calligraphic style Naskh is a piece of reed like a flute mouthpiece, shown in Figure 20. The reed's head shape corresponds to the graphic style that the qalam or feather will offer. For example, in the Naskh style the shape is beveled, and in the Maghreby style it is pointed. The feather's head in Naskh style is a flat rectangle. During the operation of writing, it has to be maintained with an inclination angle of approximately 70° with a virtual baseline.

The diacritic dot which appears as a tilted square is the typographical unit marked by the



Figure 20: Qalam of Arabic calligraphy



Figure 21: Diacritic dots positioning

(Source: Holy Quran written by the calligrapher Alhaj Hafez Mohamed Amin Alrochdi, Scrutinized and revised by General Directorate of Endowments, Baghdad, p. 43.)

{ب ت ث} = [ب]
 {ج ح خ} = [ج]
 {د ذ} = [د]
 {ر ز} = [ر]
 {س ش} = [س]
 {ص ض} = [ص]
 {ط ظ} = [ط]
 {ع غ} = [ع]
 {ب ت ث ن ي} = [ب]

Figure 22: Letter clusters

feather in use.

The system of diacritic dots plays a leading semantic role. Indeed, certain letters are characterized by the presence, number and positions of dots. The basic glyph ب gives rise to several letters according to the number of diacritic dots which appear above or below: ب, ت, and ث. Similarly, the glyph ح provokes several letters according to presence and position of the dot with respect to the basic glyph: ح, ح̇, and ح̈ (noted as [ح̈] in Figure 22). In the case of a succession of letters carrying two diacritic dots, the dots may interfere with each other. They can thus be placed horizontally one next to the other, or vertically one above the other, according to the available space above or below the basic glyph, as in Figure 21, or — even better — to stretch letters for better spacing. Moreover, according to the style, the size of the diacritic sign also changes according to the stretching size, also seen in Figure 21.

The importance of the diacritic dot goes beyond its phonetic role. This dot is also the measure unit used for regularizing the dimensions and the metrics of glyphs. Ibn Muqlah² specified letters measures in metric dots. In order to give homogeneity to let-

² Abou Ali Ben Mohamed Ben Ali bnou Muqlat [272–328 A.H./886–940 A.D.], a native of Shiraz and a minister of the Abbasid caliph in Baghdad, was one of the first theorists of Arabic calligraphy. His contribution to this art was not the invention of a new script but the application of systematic rules that determine the surface area and proportions of the individual letter-shapes with respect to one another.

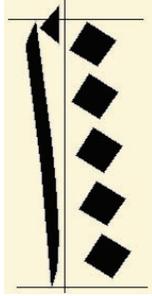


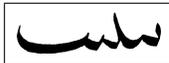
Figure 23: Arabic letter Alef metrics

ters, he includes all the Arabic letters in a circle of six dots diameter, the Alef letter height. Figure 23 shows the metric of the Arabic letter Alef in metric dot. Of course, this notion remains relative and approximate; the dot position in the line and the position of dots among them lead to differences among calligraphers.

2.6 Allographs

The allographs are the various shapes that a letter can take while keeping its place in the word: isolated, initial, median or final. Letters can have additional shapes even though, grammatically, there are only four shapes. Allometry is then the study of allographs: shape, position, context, etc. Generally, the election of an allograph responds to aesthetic needs. So, this choice is left to the writer. However, the use of an allograph is sometimes desired and even recommended. The letter's shape should change according to the letters in the neighborhood and, in some cases, according to the presence of kashida. Some examples:

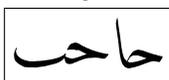
- the median form of the letter Beh should take a more acute shape if it interposes two spine letters:



- the initial form of Beh can take three allograph shapes, according to its following letter:



- the initial form of the letter Hah may take the shape of lawzi Hah whenever it precedes an ascending letter:



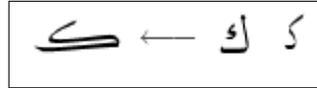
- the initial form of the letter Ain should take the shape of finjani Ain when it is followed by an ascending letter:



- the initial form of the letter Hah as well as the final form of the letter Meem change their morphological forms in the presence of kashida:



- the letter Kaf changes its morphological shape in the case of stretching and should be changed into zinadi Kaf:



3 Proposition for Arabic justification

As we have seen, to perform text justification, Arabic writing provides various techniques coming from its handwriting traditions. These techniques are not based on the insertion of variable spaces between words. So, unpleasant spaces are avoided and a balanced aspect can result.

The present study will be restricted to a formalism of the problem of text justification in the calligraphic style Naskh. For reasons of legibility, the Naskh style has been adopted in typesetting since the first attempts of computerization and standardization of Arabic typography. The Naskh style stands between the difficult Thuluth style and easy Ruq'a style. Our choice for the Naskh style is also motivated by the fact that some calligraphic styles, such as the Ruq'a style, do not allow the use of kashida. Other styles, such as the *kufi* style, have a geometrical structure which goes beyond the rules of feather and metric dot, while still other styles use letters stretching only for aesthetic purposes. On the other hand, the Nasta'liq and Diwani styles are more generous with number of kashidas: a word of three letters can indeed receive two elongations.

3.1 Breaking a paragraph into lines

Before using ligatures or kashida to justify lines, a system has to first break paragraphs into lines. A legitimate place to break a sequence of words into lines is a position in a place that gives a good length. Therefore, it may be the border of the last word of a horizontal list which generates a width equal to or less than the value of the line width. Or the border of the last word of a horizontal list which generates

a width higher than the value of the line width and, through the use of ligatures of second degree, the width of the list can be decreased to result in a value equal to or less than the line width. In general, there are several legitimate breaking places, and the badness of each possibility should be considered so that the best one can be chosen.

3.2 Elongation presence

In the following, we will focus on justification using kashidas. Making a system eager to stretch a letter leads to overcoming constraints of determination of the extensibility places with the required parameters. In this context, letter stretching obeys a set of calligraphic rules. These rules can be found in [1] and/or some Arabic treatises on calligraphy and practices.

The Arabic alphabet is composed of 28 letters:

أ ب ت ث ج ح خ د ذ
 ر ز س ش ص ط ض ظ
 ع غ ف ق ك ل م ن ه و ي

Before giving an account of the rules in detail, let's look first at some general rules of kashida:

- The first letter of a word composed only of two letters may not be stretched, but the second letter can be stretched;
- No more than one letter in a same word can be stretched;
- On a given line, generally only one word can be stretched (the albasmalah sentence in Figure 24 is an exception — there, the first word is always stretched);
- The use of the kashida is recommended where there is a succession of spine letters, to remove reading ambiguity. In particular, the letter Sin should always be followed by an elongation of two metric points;
- A word ending with the letter Yeh [ي]³ cannot be stretched. The presence of kashida in this case generates a new letter *qantara*. On the other hand, it is better to stretch a word when it ends with a pronoun: final letter Heh [ه]. In this case, the kashida should be placed before the last character.

Now, let us consider the rules concerning where to insert kashidas, their number and their degree of extensibility. According to the most general rules we

³ Brackets are used throughout to indicate a *letter family*, standing for any glyph which shares this root form, differing only in the number of dots.



Figure 24: Stretching of letter Seen in albasmalah

have found, it is advised to use the kashida in the middle of the word. In words composed of four or six letters, the kashida is to be put down on the level of the second or third letter of the word respectively. In the following, we proceed first by generalizations; we classify words according to their number of letters, and enumerate occurrences where stretching is allowed.

Words with two letters: It is strictly prohibited to stretch a word with two letters, except in the case of *سر* or *شر*. Though it represents only one letter, the initial form of the letter Seen [س] can be regarded as a succession of three glyphs: [س] [س] [س].

Words with three letters: Usually, the kashida is omitted in a word composed of three letters. Words such as *عسى* and *فتى* cannot be stretched. On the other hand, in the *albasmalah*, the word *bismi* *بسم* should always be stretched in the level of its letter Seen, as in Figure 24; this elongation was not imposed previously, as in Figure 12.

According to Imad Eddine Ibn Al'afif's⁴ opinions, one can stretch a word composed with three letters when the first letter is a ع, ط, [ج] or [س]. According to Azzettaoui,⁵ one can proceed with an elongation if the last letter is an Alef or a Lam. If we decide to stretch such a word, the kashida should be performed on the level of the second letter.

Words with four letters: Words with four letters are the most susceptible to be stretched, and it is preferable to insert kashida in the second letter. However, one should not stretch words like:

تغلب, نفير, خبير.

Words with five letters: For this case also, the treatises show differences. While Ibn Al'afif forbids the use of kashida in a word of five letters

⁴ The opinions of 'afif Eddine Mohamed Alhalabi Achirazi were a reference in the Alqalqashandi's treatise of calligraphy. One of his disciples was his son Imad Eddine Mohamed.

⁵ Abou Mohamed Ben Ahmed Azzettaoui, was born in 750 A.H. and passed away in 806 A.H. He was taught calligraphy by Chams Eddine Mohamed Ben Ali Bnou Abi Raquiba.

Precede																	Isolated	
و	ه	ن	م	ل	ك	ق	ف	[ع]	[ط]	[ص]	[س]	[ر]	[د]	[ج]	[ب]	أ		
	1	1	1	-1					+1			-1	-1	-1		1	5	[ب]
-1	-1	-1	-1	-1	-1			-1	+1			-1	-1	-1		-1		[ج]
-1	-1	2	-1	2	2	-1	-1	-1	1	2	2	1	-1	-1	2	1	3	[س][ص][ط]
	-1	-1	-1	-1				-1	+1			-1	-1	-1		-1		[ع]
-1			-1	-1				-1	+1			-1	-1	-1		1	1	ف ق
-1	-1	-1	-1	-1	-1					-1		1	1	-1		-1	3	ك
	-1	-1	-1					-1					-1	-1			3	ل
-1	-1	-1	-1	-1	-1			-1	1			1	1	-1		-1		م
	-1										-1	-1	-1		1			ه

Figure 25: Chart of allowed extensions, by context

as such word cannot be divided into two equal parts, Ibn Khalouf affirms that elongation on the second letter of the word in this case is advised and necessary.

Words with more than five letters: For this case, the general rules of letter successions in Figure 25 are the norm.

3.3 Elongation places

Some special care is necessary to examine historical calligraphic compositions from a typographical point of view. Many calligraphers went beyond some established rules. They used kashidas generously when they were remunerated by the page so that they were better paid [2].

The heavy use of elongations at the line end is the preferred method for Ibn Wahid.⁶ For Ibn Al'afif [1], elongation is best at the end of a line, it can be tolerated in the beginning, and it is forbidden in the middle.

The propositions of the treatises on the legitimate places to insert kashida show differences in some details. While Ibn Al'afif allows letter elongation in the middle of line only when absolutely necessary, Alchirazi does not have objections.

In terms of justification, the heavy use of the kashida at the end of a line by Ibn Wahid and Ibn Al'afif is due to a particular reason: calligraphers can estimate the elongation only when nearing the

limit of a line. The kashida is triggered by the distance to the end of line.

The superposition of two elongations on two consecutive lines can be seen only as a defect, even by Ibn Al'afif. To avoid the “stairs” effect resulting from such superposition, a uniform typographical grid can be advised.

3.4 Degree of extensibility

The degree of extensibility of stretchable letters depends on some contextual elements:

- the nature of the letter to stretch;
- the position of the letter in the word;
- the position of the word in the line.

In order to speed up the processing, the letters will be grouped into families. The glyphs of each family undergo the same contextual treatment as well as the same extensibility rules.

Figure 25 gives the degree of extensibility in metric dot of each letter according to its context. For example: ”The elongation of letter Beh is authorized, if it is followed by Dal, and it is prohibited, if it follows Sin”.

Let us recall that the letters {أ, [د], [ر], و} are never stretchable.

In some cases, several possibilities of treatment arise. We will focus on the most widespread and simplest ones.

Notice that:

- in boxes, each number i represents the interval $[i,12]$ in metric dots;
- an empty box represents a prohibited elongation;

⁶ Charaf Eddine Mohamed Bnou Charif Bnou Youssef Azary, known as Ibn Wahid, was born in Damascus in 647 A.H. He studied in Iraq and lived in Egypt. He was one of Yacout Almoura'simi's disciples.

- the plus sign means an allowed and approved elongation;
- the minus sign means an allowed but not approved elongation, in a number of these cases, one could use ligatures instead of kashidas.

3.5 Process

Here is our proposed method for justifying Arabic text:

- break the paragraph into lines with a specific width;
- compute the badness of the individual line;
- refer to the contextual table to determine all legitimate places to insert kashida;
- establish priorities for kashida points;
- distribute the lines' badness as parameters of kashida;
- generate curvilinear kashida with determined parameters;
- insert kashida in the selected points.

4 Conclusion

The justification in Arabic writing differs from other writing systems on at least two points. First, for centuries, hyphenation of words has not been allowed. Second, Arabic is endowed with a particular tool for justification, the kashida. If historical treatises on calligraphy provide us with a list of technical specifications, our work consists of checking the agreement among these treatises and the general use in current manuscripts, and then attempting to build a formalism of the degrees of the letters' extensibility according to their context.

Acknowledgements:

Thanks to Barbara Beeton, Karl Berry, and Khalid Sami for their editorial corrections and contributions.

References

- [1] Abi Al-abas Ahmed Ibn Ali Al-qanqachandi (821h/1418). *Sobh al-a'cha fi sna't al-incha*. Version copied from edition Al-amiriat with corrections and a detailed indexing. Ministry for the culture and national orientation, general Egyptian institution of production, edition, impression and distribution.
- [2] Vlad Atanasiu. *Le phénomène calligraphique à l'époque du sultanat mameluk*. PhD thesis, 2003. <http://www.atanasiu.freesurf.fr/thesis>.
- [3] Zeev Becker and Daniel Berry. `triroff`, an adaptation of the device-independent `troff` for formatting tri-directional text. *Electronic Publishing — Origination Dissemination and Design*, 2(3):119–142, October 1989.
- [4] Daniel M. Berry. Stretching letter and slanted-baseline formatting for Arabic, Hebrew and Persian with `ditroff/ffortid` and dynamic PostScript fonts. In *Software — Practice & Experience*, number 29:15, pages 1417–1457, 1999.
- [5] Yannis Haralambous. Tour du monde des ligatures. *Cahiers Gutenberg*, 22:69–80, 1995.
- [6] Yannis Haralambous. Voyage au centre de \TeX : composition, paragraphage, césure. *Cahiers Gutenberg*, 44-45:3–53, 2004.
- [7] Azzeddine Lazrek. CurExt, Typesetting variable-sized curved symbols. In *Euro \TeX 2003: 14th European \TeX Conference*, TUGboat 24:3, pages 323–327, Brest, France, 2003. <http://www.ucam.ac.ma/fssm/rydarab/doc/communic/curext.pdf>.
- [8] Liang, Franklin Mark. *Word Hy-phen-a-tion by Com-put-er*. PhD thesis, 1983. <http://tug.org/docs/liang>.
- [9] Thomas Milo. ALI-BABA and the 40 Unicode Character — Towards the Ideal Arabic Working Environment, New input output concepts under Unicode. In *Euro \TeX 2003: 14th European \TeX Conference*, TUGboat 24:3, pages 502–511, Brest, France, 2003.
- [10] Johny Srouji and Daniel M. Berry. Arabic formatting with `ditroff/ffortid`. In *Electronic Publishing — Origination Dissemination and Design*, volume 5, pages 163–208, 1992.
- [11] Hàn Thế Thành. Améliorer la typographie de \TeX . *Cahiers Gutenberg, actes du congrès GUT'99*, 32, 1999.
- [12] Adolf Wild. La typographie de la Bible de Gutenberg. *Cahiers Gutenberg*, 22:5–15, 1995.