

# Allographic Biometrics and Behavior Synthesis

Vlad Atanasiu

Massachusetts Institute of Technology

USA

atanasiu@mit.edu

<http://mywebpage.netscape.com/atanasiuvlad/frq/>

## Abstract

This essay deals with the role of allographs in the general framework of the study of writing. It explains why allographs are important (writer individuality), how they can be measured (frequency and contexts (graphic, semantic, non-graphic)) and what applications to expect from them (biometrics, script restoration, artistic creation, cultural history).

## Résumé

Cet essai traite du rôle des allographes dans le cadre général de l'étude de l'écriture. Il explique pourquoi les allographes sont importants (l'individualité de l'écrivain), comment les mesurer (fréquences et contextes graphiques, sémantiques et non-graphiques) et à quelles applications s'attendre (biométrie, restauration de l'écriture, création artistique, histoire culturelle).

One of the common elements used to identify a typeface or a handwriting is the shape of the script. Sometimes, however, individual characters may have alternate forms — in those cases the rules and habits governing their usage characterize the behavior of the typesetter or writer in a more subtle way than shape does. Allographs provide inconspicuous signatures to a text.

This paper intends to raise the issue of the study of allographs — allometry —, for the pleasure of its intricacies as well as for its many practical applications. I will proceed by discussing four diagrams, representing some of the central issues of the subject. Although my discussion is based on Arabic handwritten examples — a script where allographs are an inherent part of the writing system —, the same principles can be applied to non-Arabic hand- or machine-produced scripts.

### *Measuring frequencies*

The two text blocks in Fig. 1 represent a double-page of a manuscript copied in Sicily in the 14<sup>th</sup> century [1]. On the right side we see the original script (Fig. 1a), whereas the left side (Fig. 1b) is a replacement folio, added at a later date, after some folios went missing. Some attempts were made to keep the aspect even: the dimension of the text-block and the line number is the same and both scripts are in the *naskh* style. But even an eye unaccustomed to Arabic script will feel the presence of crude restoration work.

The paleographic expertise can be automated; the magnitude of a 2D Fast Fourier Transform of the pages will reveal two main differences. The axis-angle of the cross-like spectra are different, which means that the ver-

tical strokes of the replacement script are more slanted than the original ones (Fig. 1c has a nearly horizontal axis, Fig. 1d a more oblique one; the vertical axis comes from the regular spacing of the 13 lines). The four bright spots away from the center of Fig. 1d suggest a stronger linearity of the replacement script, whereas the dispersion in Fig. 1c points to a rounder and more irregular handwriting in the original script).

But there is an even more powerful distinguishing feature: two allographs of the *kaf* letter are used by the replacement, the oblique one (marked by a circle, 13 times in Fig. 1f) and the long one (marked with a disc, 7 times in Fig. 1f), whereas only the oblique appears in the original (19 times in Fig. 1e). It is also remarkable that most (4) come in the context of the same word (*kana* “to be”). Defined by quantities and contexts never obvious to the eye, the allographs offer a reliable security filigree for testing the integrity of the manuscript.

### *Applications*

Figs. 2a and 2b are excellent examples of the principal application of allometry: *biometrics*, or the spatio-temporal identification of scripts and their attribution to individuals or traditions (a subject of interest to police and intelligence agencies, the judicial system, the art market, historians, et al.). The examples presented here are written by two calligraphers for a competition and transcribe the same text [2, p. 31–32]; however, the number of long *kaf*-s is 6 and 23, respectively. Thus the allographs of the elongated type clearly play a very different role as “visual attractors” from one individual to

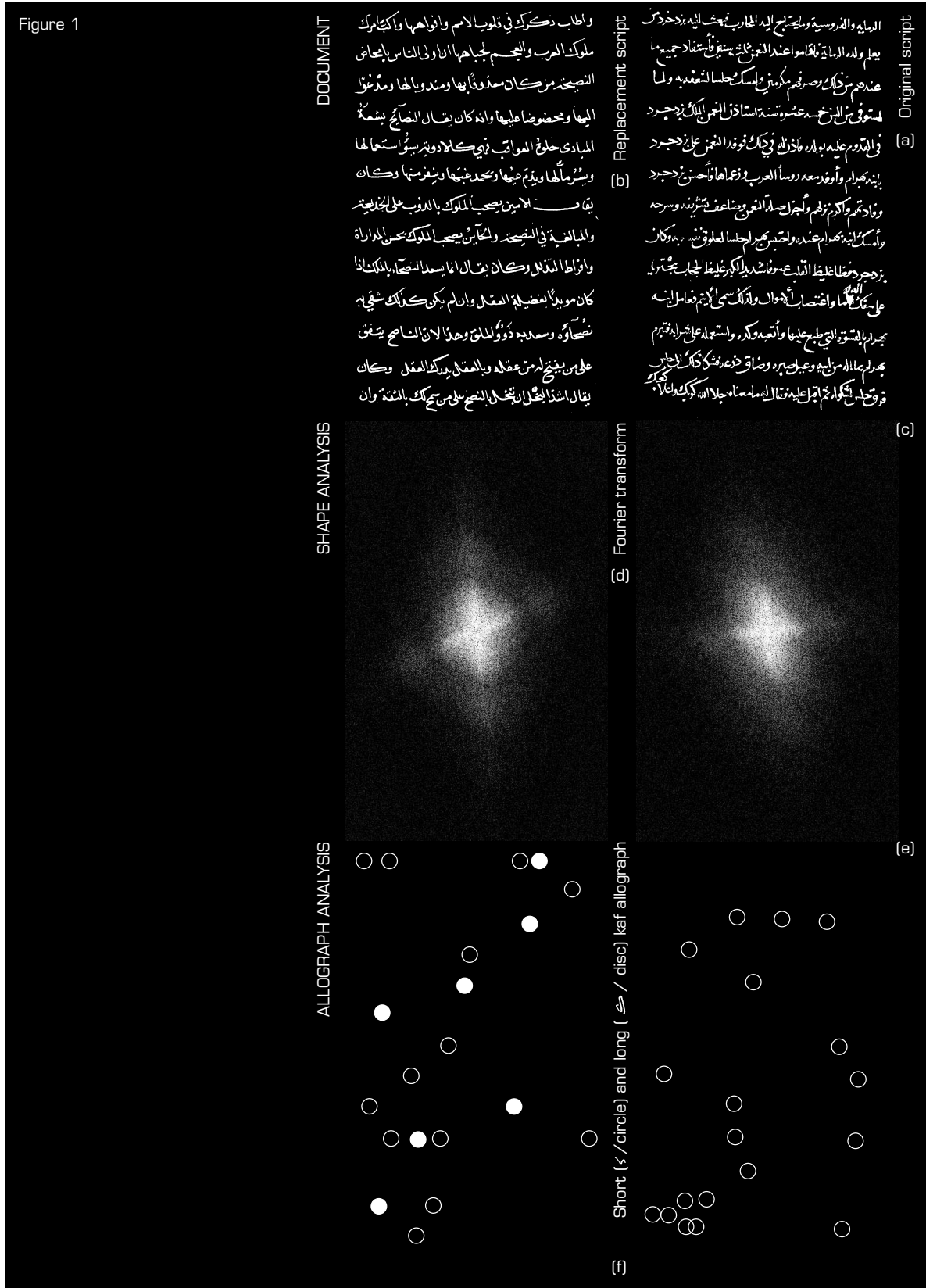


FIG. 1: Letters beyond shapes: quantity and contexts

يوطن ونسك حركي . ولا  
 يُعزلنا عنكم بغيركم إلا .  
 فصلهم وهم مؤمنون بالحق وضربوا عنقاً غير مسيح . فإن آمنوا وأطعتم  
 وكفروا بالمرء . وإنما النساء عنكم عريان لا يلبسهن شيئا . أخذوا من آيات الله  
 واستحلوا فرجهم بكمية آفة . فأقروا الله عيب النساء . واستنصوا بهن عيرا . الأهل بلغت ؟ أهنا شهدا  
 إنما الناس . إنما المؤمنون إخوة . وأخيل لا يهتد مال الشيرا إلا عن طيب قسرينه . الأهل بلغت ؟ اللهم شهدا  
 فلا تخبرين عدي كذا يا بصرى . بفسك رفات بغير . فإني قد تكبر . إن أشقتم يولد  
 خيرا منكم . كآفة . الأهل بلغت ؟ اللهم شهدا .  
 إنما الناس . إننا نأكل وأما . إن لآدم . وآدم من ذرية . أكرهت آفة  
 أفتاك . وليركض على عصى . الأهل بلغت ؟ اللهم شهدا . فإنا  
 فليبلغ الشاهد الشاهد .  
 إنما الناس . إن آفة قد نكحوا وارتب . حبيبه من المرات . ولا يجوز وصية . ولا يجوز وصية  
 في الكرم من الفس . وأولاد الفساش . ولت جبر الجبر . مرادع الجبرانية . أو قول غير تواليه . فليبه  
 فآفة الله وللأكل والناس الجبين . لأيقنا به ضررت ولا عمل . والسلام عليكم ورحمة آفة .

يوطن ونسك حركي . ولا  
 يُعزلنا عنكم بغيركم إلا .  
 فصلهم وهم مؤمنون بالحق وضربوا عنقاً غير مسيح . فإن آمنوا وأطعتم  
 وكفروا بالمرء . وإنما النساء عنكم عريان لا يلبسهن شيئا . أخذوا من آيات الله  
 واستحلوا فرجهم بكمية آفة . فأقروا الله عيب النساء . واستنصوا بهن عيرا . الأهل بلغت ؟ أهنا شهدا  
 إنما الناس . إنما المؤمنون إخوة . وأخيل لا يهتد مال الشيرا إلا عن طيب قسرينه . الأهل بلغت ؟ اللهم شهدا  
 فلا تخبرين عدي كذا يا بصرى . بفسك رفات بغير . فإني قد تكبر . إن أشقتم يولد  
 خيرا منكم . كآفة . الأهل بلغت ؟ اللهم شهدا .  
 إنما الناس . إننا نأكل وأما . إن لآدم . وآدم من ذرية . أكرهت آفة  
 أفتاك . وليركض على عصى . الأهل بلغت ؟ اللهم شهدا . فإنا  
 فليبلغ الشاهد الشاهد .  
 إنما الناس . إن آفة قد نكحوا وارتب . حبيبه من المرات . ولا يجوز وصية . ولا يجوز وصية  
 في الكرم من الفس . وأولاد الفساش . ولت جبر الجبر . مرادع الجبرانية . أو قول غير تواليه . فليبه  
 فآفة الله وللأكل والناس الجبين . لأيقنا به ضررت ولا عمل . والسلام عليكم ورحمة آفة .

Figure 2

(b)

(a)

Strong and weak visual attractors:

for the same text, but for different calligraphers the usage of the rare long <kaf> allograph is 6 for 23.

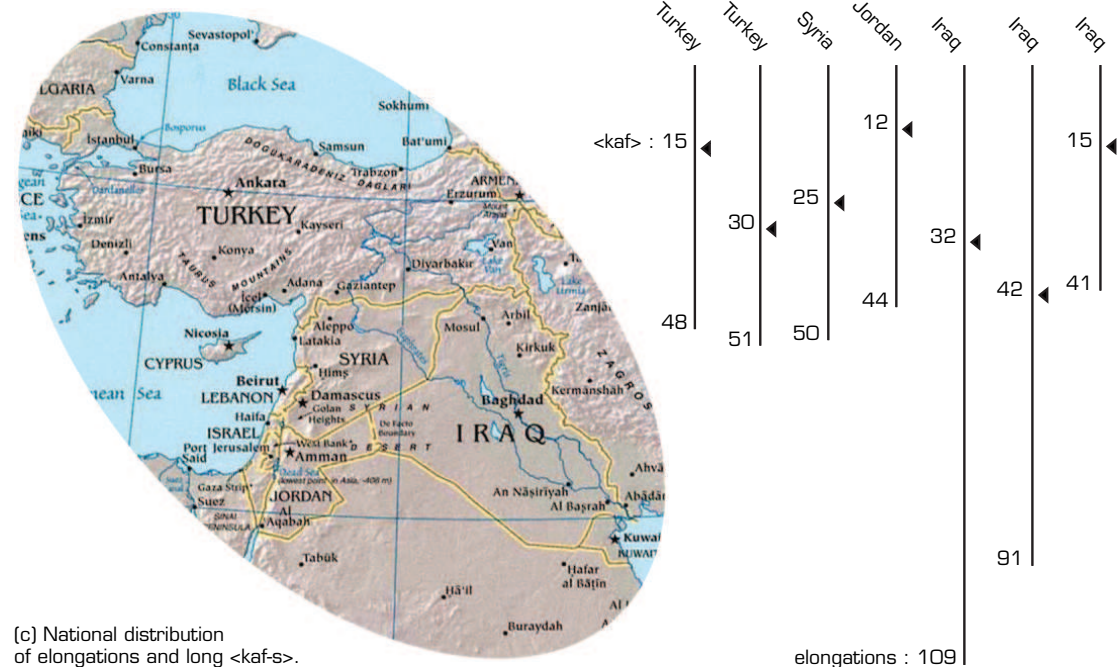


FIG. 2: Allographic behaviours: individuals and traditions

another. The same can be observed at the level of traditions. On the basis of five other participants in the calligraphic competition, we see that Turks globally use a very similar number of elongations (but differ individually in their attribution to specific graphemes). They are closely connected to the Syrian, Jordanian and part of the Iraqi trends. In Iraq, however, two traditions prevail, with one making an extensive use of elongations, very unlike Turkish calligraphers.

Once the script analysis is done, it can serve as a basis for a *cultural history*. Thus, the coherence of the Turkish style is remarkable, witness to the force still extant today of a tradition hundreds of years old. The allographic differences are the expression of an Arabic style independent of the Turkish school — an anti-monopolistic tendency anchored in the political, economic and social make-up of the Middle East. Given that the Iranian calligraphic styles need a good deal of elongation, one also has to ask what the influence of Iran is on its neighboring Iraqi calligraphy. Arabic calligraphy still being an important artistic player in Islamic countries, such analysis can reveal some preferential channels of influence and provide clues for policymaking.

Another active use of allometry is *behavioral synthesis*; after having extracted information about the way a person or tradition handles allographs, it can be used to predict future behavior, whether in known textual situations (typically the restoration of a damaged or missing part of a writing) or for virtual texts, never written by the individual (these are revival texts, “in the style of ...”). Models of behavior should be seen as an important aspect of typographic automation for its capacity to adapt graphic output to specific contents or lectorates (an Ottoman text with an Arabic allograph frequency would be historically odd and certainly disapproved of by the calligraphic establishment).

Finally, allometry can be used to enhance the visual aspect of a text in ways too complex or tedious to be feasible by human minds and hands. For example, allographs are known to be used for producing graphic figures or rhythms, but they occur almost always locally, at the page level, and stretch very rarely across a manuscript of several hundred pages (which thus becomes more a suite of chamber music than a symphony). However, with a program for *computer aided script*, a designer could concentrate on modeling the figures and rhythms, fine tuning and harmonizing them, and leaving the detail rendering to the machine. Just as sound synthesizers help composers, in the same way could computers generate written texts of a far more complex structure than anything seen before.

At a methodological level, allography leads us to define the intellectual goals of the study of writing (*grapho-*

*nomics*<sup>1</sup>) as follows. The graphic *analysis* is a required step of script identification. It enables the building of a *model* for the reproduction of a specific writing behaviour and in a later stage its modification and use in the process of new graphic *creations*. Script recognition leads also to the understanding of the non-graphical aspects that shaped its existence: Beside its *physical aspects* (spatial, temporal, force, acoustics, ...), this graphical ecology is *technological* (materials, tools, process), *informational* (mathematical, statistical, informational properties of writing systems), *biological* (physiological, neurological, psychological) and *cultural* (social, economic, artistic, literary, customary, religious, ..., aspects of writing).

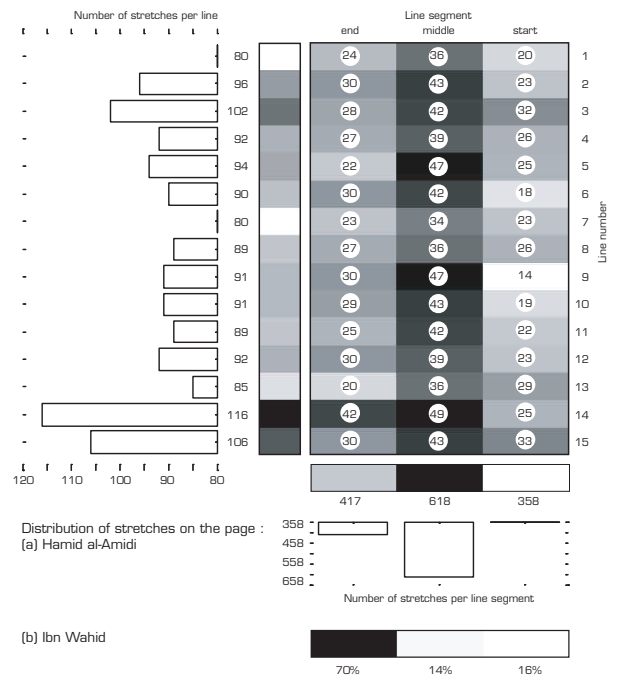


FIG. 3: The simultaneous contexts problem

*Measuring semantic contexts*

In a calligrapher’s mind the presence of a word with a strong meaning (for himself or within the context of the text) implies using a special graphic output for that word. In an Arabic working text this happens generally by using less frequent allographs. In Fig. 3 we see the example of a word containing the letter *kaf*, *kufr*, which means “impiety”. In the religious context of the text — Qur’anic citations, holy sayings, moral advice — this word triggers in 70% of cases the usage of an alternate allograph in the group of calligraphers studied (the same as for Fig. 2).

1. Graphonomics is a field of study different from graphology, in the same way that astronomy is distinct from astrology. It has its own association, conference and publication; see <http://www.cedar.buffalo.edu/igs>.

In contrast, *malaka* “to possess”, a word less relevant to the text, is only in 20% of the cases attributed the long, exceptional, *kaf*. A side effect of the transfer from lexical meaning to visual display of exceptional status, is the creation of a cultural bond between the otherwise random association of a meaning and a specific shape. Poets have started to use the expression *kaf-i kufr* “the *kaf* of impiety” and explain the elongated shape of the letter by comparing it to the “narrowness of heart” of some people. The metaphor strengthens the practice of the calligrapher by wrapping it in a cultural halo and a web of meanings. The semantic triggering of an allograph exemplifies the second most important aspect of an allograph, after its frequency: the contexts where and why it is used.<sup>2</sup>

### Problems and limits

Fig. 4 presents a case of space management — that is, the various techniques for fitting a specified amount of text on a finite inscribable surface. Here I analyze the behavior of a writer regarding the elongation of graphic segments in the particular setting of the Qur’ans of the Ottoman tradition called “*ayet ber kenar*”, where a verse ending must coincide with the end of a page, thus forcing the calligrapher to carefully consider not only the line, but also the page justification (more references to this calligraphic feat in [4]). The superposition of all elongations of part of a manuscript [5, p. 2–48] show a fluctuation in the quantity of elongations from line to line. This means that the text was not studied so as to let the elongations spread uniformly across the page and that the writer needs an adjustment period observed in the first third of the page before settling to a more constant quantity and then again overusing the elongations during the two lines before last to fit his text. He shows, however, a stronger ability of prescience — or faithfulness in his credo of uniform justification — than the Mamluk calligrapher of Fig. 4b, who comes to the conclusion that he needs an elongation in order to justify only at the very end of lines [6, vol. 3, fol. 99v<sup>o</sup>–166v<sup>o</sup>]. So the Turkish calligrapher starts thinking ahead earlier, generally in the middle of a line. The low quantity of elongations in the first part of lines indicate that they are primarily used as justification devices. Technical treatises of Arabic calligraphy and typography confirm these findings.

However, if we want to quantify the justification context, we have to consider that what we see on a page is the result of the simultaneous presence of several contexts, each having specific demands on the graphic out-

put. For example, we saw in Fig. 1 that the long *kaf* is an important part of all elongations, thus the uneven distribution of elongations could be partly due to a particular distribution of words containing the letter *kaf*. In fact it is not possible to consider just one context at a time, and a theoretic framework has to be developed to disentangle the contribution of each context.

The principal hurdle comes from the number of potential contexts that can coexist (consider also other aesthetic reasons that could have led the calligrapher to make the choices he did, such as line or word balance, black vs. white equilibrium, etc.), from the “melting” of contexts (how to know which ones actively influenced the choice?) and the knowledge that we can have about them (some are “hidden” — such as page justification would probably be for a casual reader —, yet others are “forgotten” — such as the mood of the writer, which certainly was a factor in his performance). These are the frontiers of our present knowledge — some to be left unchallenged until time machines and nano-submarines for brain exploration are invented.

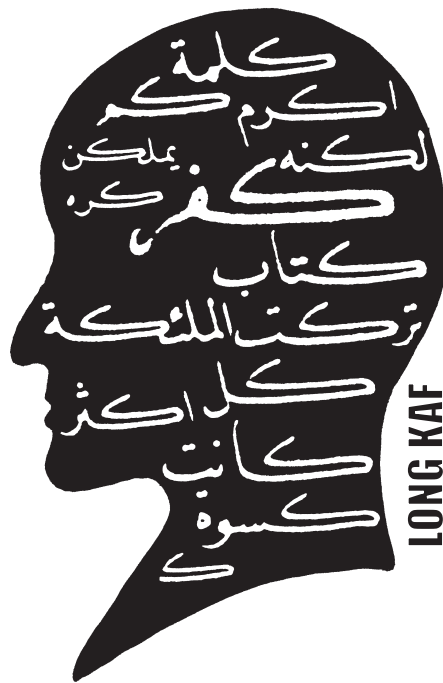
### References

- [1] Ibn Zafir, *Sulwan al-muta’ fi ‘udwan al-ittiba’*, TRI Publishing, Kuwait, 1984.
- [2] Mohammed Tamimi, *Catalogue of Winners’ Plates in the Second International Calligraphy Competition*, Research Centre for Islamic History, Art and Culture, Istanbul, 1991.
- [3] Vlad Atanasiu, *Hypercalligraphie. Le phénomène mamluk à l’époque du sultanat mamluk (Moyen Orient, XIII<sup>e</sup>-XVI<sup>e</sup> siècle)*, thèse de doctorat, École pratique des Hautes Études, Paris, 2003, <http://mywebpage.netscape.com/atanasiuvlad/the/>.
- [4] —, “Visible *i’ajaz* — Three graphic particularities in the Qur’an copies”, in *Actes du colloque de Bologne 2003 sur les Manuscrits coraniques*, François Déroche ed., E.J. Brill, Leiden, 2003, <http://mywebpage.netscape.com/atanasiuvlad/pubs/atanasiu2003iajaz.pdf>.
- [5] *Qur’an*, calligraphed by Hamid al-Amidi, Hizmet Vakfi, Istanbul, 1974.
- [6] Ibn Wahid, *Qur’an*, Cairo, 1304–6 (London, British Library, ms. Or. Add. 22408).

2. A classification of allographic contexts for the Arabic script and its comparison with Latin script can be found in my doctoral thesis [3, chap. 10, “La grammaire graphique”].

**YOU  
CLOTH  
BUT  
GENEROSITY  
MOST  
ANGELS  
WORD** 40%

**LEAVE  
EVERY  
BOOK  
BEING** 50%



**LIKE  
POSSESS  
HATE** 20%

**IMPIETY** 70%

FIG. 4: Contextual aspects: association strength between meaning and shape