Le Stratifiant : a simple tool for processing stratigraphic data*

Bruno DESACHY
Université Paris-1, France

Abstract: This paper presents an application (Le Stratifiant) which makes it possible to create automatically stratigraphic diagrams. Le Stratifiant uses a simple algorithm, based on an adjacency matrix. It provides a graphic representation slightly different from the Harris Matrix, in order to avoid ambiguities, and to represent more easily different types of relationship. It can also processes indicators of dating (terminus post quem, terminus ante quem), in order to assign a bracket of dating to each stratigraphic unit, and to visualise these brackets. Another important characteristic, which is also a methodological choice, is the possibility of processing two levels of relations: normal, or uncertain.

Le Stratifiant is developed on a preexisting software (Microsoft Excel). This choice allows to benefit from all the possibilities of this common software, like import – exports files and communication with data bases.

Keywords: archaeological stratigraphy; Harris Matrix computerized application

Introduction
The way chosen for the application presented here is to exploit the possibilities of one of the most known softwares in the world, set up in most computers - Microsoft Excel - , in order to create a simple tool for processing stratigraphic data. Excel is a very common spreadsheet ; but it is also a development tool, which allows to create specific applications, without acquisition of a new software, and remaining in the same well-known environment of work. The tool presented here, called Le Stratifiant, comes as an Excel file. It may be moved and copied like any other Excel file. But when it is opened, some specific characteristics appear : it contains a program written in VBA (the Microsoft application programming language), preset sheets to receive the data, and buttons and controls which allows to operate the data processing exposed below.

* This paper has been presented at the 13th Congress in 2008
Fig. 1 – Data sheet and options window in Le Stratifiant

**Formal view of stratigraphic relations**

All the existing stratigraphic data processing tools, *Le Stratifiant* included, need to represent stratigraphic relations by mathematical relations with defined properties. Mathematically, anteriority ("prior to") and posteriority ("after") relations are reciprocal partial orders (orton 1980), according to their properties:

- irreflexivity (a layer can't be prior to itself);
- transitivity (if a layer is prior to another layer, and if this second layer is prior to a third layer, then the first one is prior to the third one);
- antisymmetry (if, like Clive Orton, we regard stratigraphic relations as a non-strict partial orders, so that if a layer A is prior to a layer B, and symmetrically if B is prior to A, then A and B are the same layer), or asymmetry (if we regard stratigraphic relations as strict partial orders, so that if A is prior to B, then B can't be prior to A).

As representation of a partial order, the "Harris matrix" comes under the graph theory: it is formally an oriented graph, made of nodes (the stratigraphic units) and arcs (the lines joining the units). Ancestors of the Harris matrix may be found in some specific applications of graph theory: the tasks scheduling graphs, perfected in the years 1950 to manage complex industrial and military production processes. The most known of these ones is PERT (Program (or project) Evaluation and Review Technique) chart, created in 1957 to manage the development of the US Polaris strategic missile. In a PERT graph, events, which are typically tasks of a production process (for instance “set windscreen” in a cars assembly line), are represented by arcs. The node between two successive arcs represent the “precedence constraint” between the two tasks: (for instance, to build a car, « set windscreen » and « windscreen wipers »). Another tasks scheduling technique is the MPM (Méthode des Potentiels Metra) graph, designed since 1959 by the mathematician Bernard Roy (roy 1959, faure lemaire picouleau 2000). In this graph, nodes are tasks, and arcs are precedence constraints. The MPM graph is more in accordance with graph theory and so it is more efficient (it doesn’t need to insert a “dummy task” to complete the graph when a task has several prior events).

It is easy to see how the tasks scheduling problems are related to the archaeological stratigraphic problems: a stratigraphic unit may be regarded as a “fossil event” in a site formation process; and the stratigraphic relations are the precedence constraints between these “fossil events”. From this point of view, the MPM
graph is formally very close to the Harris matrix (DESACHY DJINDJIAN 1990). So, the process exposed below is greatly inspired by these tasks scheduling techniques and particularly by the MPM technique.

![Diagram](image)

Fig. 2 – representations of a partially ordered sequence of five events: a) Harris Matrix, b) MPM graph, c) PERT chart. In the two former cases, the events are represented by the nodes. In the third case, the events are represented by arcs (lines); a “dummy” (non existing) event must be added to represent the complete precedence constraints.

Archaeologists record not only anterority or posteriority relations, but also stratigraphic synchronisms. As we can consider antéro-posteriority to be a mathematical partial order, we can theoretically regard those stratigraphic synchronisms as a mathematical equivalence, with its properties of reflexivity (a layer is synchronous with itself), transitivity (if a layer is synchronous with an another one, and if this second layer is synchronous with a third layer, then the first one is synchronous with the third one), and symmetry (if a layer A is synchronous with a layer B, then B is synchronous with A). However, in some cases, this mathematical relation of equivalence is not sufficient to take in account the whole notion of stratigraphic synchronism (or “correlation”) practically used by the archeologists (what we will discuss below).

**Stratigraphic data processing**

Like tasks scheduling graphs and other oriented graphs, Stratigraphic diagram can be built with two sets: the set of nodes and the set of oriented pairs of nodes which define the arcs. The archaeological observations provide these two sets, which are the list of the units, and the list of the elementary recorded relations (pairs of units). These two lists must be entered in *Le Stratifiant*, in the appropriate sheet (Fig. 1). The kind of relation must be indicated: antero-posteriority, or synchronism. The relations must be entered in *Le Stratifiant* in the way of anteriority (the first unit is the anterior one; the second is the later one). These lists may be directly keyboarded, or copied and pasted, or imported from a data base.

The process is based on an adjacency matrix. This data structure, used by some authors (DESACHY DJINDJIAN 1990, SHARON 1996), has a disadvantage (HERZOG 1993): the storage requirement increase exponentially with the number of stratigraphic units. But it has significant advantages: a great simplicity of process, just by scanning the matrix; and all the relations can be stored in it, even the logically impossible ones, so the detection of the recording faults is easy.

Starting from the lists of recorded units and relations, the matrix is coded with the units in rows and columns (by convention, the posterior units are the columns; the former units are the rows). The order recorded relations are coded in the squares (value 1).

Then, these matrix order values are completed and recalculated, and all the purely logical relations (not materialized in interfaces) are transitively deduced, by scanning the matrix, using the formula:

\[
\text{if } \text{matrix}(i, j) \neq 0 \text{ and if } \text{matrix}(j,k) \neq 0, \text{ then } \text{matrix}(i,k) = \text{matrix}(i,j) + \text{matrix}(j,k).
\]
Several iterations may be necessary. At the end of the calculation, for each pair of units, the obtained value is the stratigraphic distance, *i.e.* the number of stratigraphic steps between the two units. If a unit is directly above another one, the stratigraphic distance remains 1. "Redundant" (transitively deductible) relations have distances higher than 1. A value 0 means that the two units are not in relations. Then, for each unit, all its former units appear in its column with their order given by the stratigraphic distance; and in the same way, all its posterior units appear in its row.

Each unit has a distance from the oldest, *i.e.* the beginning of the sequence (the higher value in its column), and a distance from the most recent, *i.e.* the end of the sequence (the higher value in its row). These distances from the end and the beginning of the sequence will define the stratigraphic rank and the height of the unit in the diagram.

![Simple example of stratification](HARRIS 1979: Fig. 28)

![Adjacency matrix](visible order relation coded)

![Processed adjacency matrix](stratigraphic distances)

**Graph drawing**

The graphs produced by *Le Stratifiant* are quite different from the traditional Harris diagrams (and from the graphic form of MPM or PERT graphs). Relations are always represented by right lines, never by broken lines. The orientation of lines has a strict significance. Lines are either vertical, or horizontal. Vertical lines represent relations of stratigraphic order (anteriorty, posteriority). Horizontal lines represent relations of stratigraphic equivalence (synchronism).

To respect these conventions, if a unit has several non-redundant former or posterior ones, it is necessary to repeat this unit; in synchronic (equivalence) relation with itself (this reflexive equivalence relation is marked by a double horizontal line).

This graphic form has been chosen because it has a clear graphic logic which avoids ambiguities (some ambiguities are possible with the Harris classical form: BIBBY 1993, ROSKAMS 2001), and it is easy to build (even by hand).

![Harris and Le Stratifiant representations](Fig. 4 - Harris and Le Stratifiant representations).
The diagram is drawn, path after path, by scanning the processed adjacency matrix, using only values 1 (no redundant relations), are not drawn. So, Le Stratifiant respects a basic principle of the Harris matrix (known as the "fourth law" or Harris): showing all the chronological sequence deduced from the stratigraphy, but only this relative chronological sequence (and not the whole physical stratigraphy).

Fig. 5 – each path of the diagram is drawn by scanning the matrix, using the values 1 (non redundant relations).

The order of scanning matrix influences the shape of the diagram. At this stage, there are some instructions in order to reduce crossings between horizontal and vertical lines (in other words, in order to improve the planarity of the graph). Anyway, with the form of diagram produced by Le Stratifiant, these crossings are not so annoying. Indeed, it's impossible to confuse them with relations, because relations are never represented by broken lines. At the end of the process, the redundant relations, not displayed on the diagram, are displayed in gray in the list of the relations (on the data sheet showed Fig. 1).

**Synchronisms and uncertain relations**

The main specific characteristic of Le Stratifiant is to manage two levels of relation: certain relations, and dubious or uncertain relations.

This distinction is useful, first, to deal with the relations of synchronism. It allows to clarify some ambiguities which often stick to these relations. In the example shown in the figures, Harris indicates what he calls a "correlation" between the two units 7 and 8 (supposed to be the same unit at the beginning). There are two ways to consider this “correlation”.

On one hand, we can estimate that this common origin of the two layers is certain. In this case, layers 7 and 8 are indeed the same stratigraphic unit. So, this correlation is a real synchronism; in other words, a relation of stratigraphic equivalence. So, following the rules fixed by Harris himself, the redundant relations must be removed. The representation of the stratigraphic sequence is thus simplified (Fig. 6a). Practically, the synchronous units, identified by a specific coding value of synchronisms in the adjacency matrix, are merged during the process, before treating the order relations.

On the other hand, we can estimate that the common origin of these two layers is only an assumption of the archaeologist, not a certain fact. The example of Harris seems to come under this case of uncertain relation: as Steve Roskams said (ROSKAMS 2001), this correlation is not based on a direct observation of a physical continuity, but only on an extrapolation. In this case, the two units are placed at the same height and are connected visually in the diagram, but they are not merged during the process (Fig 6b). The rule of elimination of the redundant relations does not apply.
If we consider these two levels of relation (certain and uncertain), we think that the relation of certain synchronism is to be reserved when a continuity or intrication between two units is directly observed, so that they constitute, with no doubts, only one stratigraphic step. Such synchronic units may be distinguished because of their different nature (for example a buried corpse and the funerary deposit which accompanies it) or simply for technical reasons, (for instance a wall, renumbered in two different sectors of excavation). Anyway, it is the responsibility of the archaeologist to define the level of its observation, as certain or dubious. *Le Stratifiant* processes the data accordingly.

These two levels may, and logically must, be also applicable to the anteriority and posteriority relations. They allow to manage cases of uncertainty of stratigraphic succession; which are possible, for instance in the recording of standing structures. Practically, the uncertain order relations are represented by vertical dotted lines.

The uncertain character of an order relation results in a restriction of the rule of elimination of the redundant ones. This rule is segmented: a certain relation is regarded as redundant (thus eliminated from the diagram), only compared to the relations of the same mode (certain). So, a certain relation, redundant only compared to an uncertain one, will be drawn on the diagram because it shows the continuity of the certain chronology. Uncertain relations are identified by negative values in the adjacency matrix (in this case, the stratigraphic distance is calculated with the absolute value).

---

**Fig. 6** – certain synchronism (with elimination of redundant relations) and uncertain synchronism with *Le Stratifiant*.

**Fig. 7** – Uncertain order relation between two pits: the archaeologist thinks the pit (104, 102) cuts the pit (105, 103), but he is not sure. This uncertain relation is specifically displayed on the diagram (dotted line), with "redundant" certain relations to keep the continuity of the certain chronology (a). If other searchers are not convinced by the archaeologist arguments about this dubious relation, an another chronology may be suggested; but the certain relations don’t change (b).
From stratigraphic data to quantified time

Le Stratifiant offers a function of inscription of the stratigraphic sequence in the quantified absolute time, if at least few units have known indicators of dating, terminus post quem (TPQ) or terminus ante quem (TAQ), provided by historical evidences or post-exavation analysis.

Before presenting this function, it is necessary to remind, or make clearer, some points about the relationships between the ordered stratigraphic time and the quantified absolute time.

Into the quantified time, a stratigraphic unit has duration of formation, which is a length of time, with a beginning and an end. Like a precedence constraint in a PERT or MPM graph, an elementary stratigraphic relation between two units may itself represent a length of time: the time between the end of formation of the anterior unit, and the beginning of formation of the later one.

From this point of view, we must notice that a Harris matrix gives a defined – and so, limited – information; it represents each unit at a specific point in the quantified time: the end of formation. The diagram only gives the succession of the moments of end of formation of the units; it tells nothing about other chronological information. For instance, it tells nothing about formation durations of the units. The Harris matrix is a useful tool, but not a universal chronological tool.

However, for the archeologists, this moment of end of formation of a unit is important, because it is this moment which is limited by the common indicators of dating: TPQ (after which the end of the formation of the unit is necessary placed) and TAQ (before which the end of the formation of the unit is necessary placed). In fact, this specific point in the time – the end of formation of the unit – corresponds with what is generally understood by “date” of the unit (although this notion of date of an archaeological layer may be sometimes ambiguous).

The relation between the date of a unit (according to the definition above: the final moment of formation) and its TPQ is a partial order (equal or later than), like the stratigraphic relations. So TPQ and TAQ can be transferred by the way of stratigraphic relations and their property of transitivity. If a unit $i$ is prior to a unit $j$, and if the TPQ of $i$ is the year $y$, then we can write:

$$y =< date(i) < date(j) ; \text{ so : } y =< date(j) ;$$

And vice-versa, if the TAQ of $j$ is the year $z$:

$$date(i) < date(j) =< z ; \text{ so : } date(i) =< z$$

These transfers are usually practised, but often empirically and incompletely. Le Stratifiant allows to systematize them, for the whole recorded stratification.

We can remark that the other moments and lengths in time mentioned above – the lengths of formation of units, the moments of beginning of formation, the lengths of time represented by the relations themselves – may sometimes be estimated with intervals, like the TPQ - TAQ intervals. Then, they could be integrated in these systems of inequations based on the stratigraphic relations (DESACHY 2008). Processing such completed systems of inequations is the next step of development of Le Stratifiant; at present it processes only TPQ-TAQ indicators.

Practically, it is necessary to choose an "absolute" TPQ and an "absolute" TAQ, valid with no doubt for the whole recorded stratification (for example the date of the beginning of the excavation for TAQ). At the beginning of the process, all the units have this interval.
If a layer has a specific TPQ (provided for instance by a datable artifact, or by a radiocarbon analysis...), this one is transferred via stratigraphy on the posterior layers (except for the posteriors units which have more recent TPQ); and vice versa, each specific TAQ is transferred on the prior units ; so that the TPQ-TAQ interval of each unit is recalculated.

After these transfers, the units may be sorted into TPQ order, before drawing the diagram. Thresholds corresponding to the TPQ are drawn in the diagram (red horizontal dotted lines); so that the first threshold line under a unit is its TPQ. In other words, the diagram has a layout fixed at the earliest in the quantified time.

The other choice is sorting the units into TAQ order. Then, the process leads to a diagram fixed on the TAQ: so that the first threshold line above a unit is its TAQ. In other words the diagram has a layout fixed at the latest.

So, starting from a recorded stratification with some TPQ and TAQ, Le Stratifiant can provide not just one diagram, but two: one fixed on the TPQ (Fig. 8a) and other one fixed on the TAQ (Fig. 8b). It is not possible to represent both TPQ and TAQ by horizontal thresholds lines on the same diagram, because of the disparity of the intervals TPQ-TAQ (in other words, because the imprecision of dating is not constant).

These two diagrams are the two extreme shapes that the diagram can take according to the chronological quantified indicators. The archaeologist may choose, between these two extremes, one or different intermediary layouts. For that, he can modify the diagram obtained automatically, as we will see below.

To represent easily the TPQ – TAQ intervals, it is necessary to use a chart specifically based on a quantified scale of time. This TPQ – TAQ intervals chart is the third document resulting from the process, complementary to the diagrams (Fig. 8c).

Le Stratifiant has also a function of assistance to phasing, starting from indications of phase given by the archaeologist for at least a part of the units (Fig. 11). If the user hesitates to allot a unit to a phase or to another one, he can allot a phase at the earliest and a phase at the latest. Then Le Stratifiant provides a general phasing either at the earliest, or at the latest.

TPQ and TAQ can be allotted to the phases. These phases TPQ and TAQ are then also used to calculate the TPQ-TAQ interval of each unit.
Detection of recording logical faults

Another important characteristic of *Le Stratifiant* is to provide detailed information on the logical faults of recording, in order to help the user to control and manage better his recording. However, some faults of stratigraphic recording are not detectable by an automatic device, because they do not involve inconsistencies. It is possible to detect only the faults which involve illogical result, generating cycles in the stratigraphic sequence. With *Le Stratifiant*, those recording logical faults are detected by appearance of symmetrical and reflexive relations (resulting in values other than 0 in the diagonal of the adjacency matrix), during the order relations process. It allows to identify the cycles, with the involved units and relations. In this case, the process is stopped. Units and relations involved in cycle are marked on the data sheet. Then, it is the responsibility of the archaeologist to check the recorded data and to correct them. In the worst case, if the error cannot be precisely localized, the two ways of dealing with the problem are: a) eliminating all the relations involved in the cycle, or b) to amalgamate in only one unit all the layer implied in a cycle (by application of the property of antisymmetry). The cycle should not be cut at random, because according to the place of the cut, it may be obtained different logical sequences, but each other contradictory.

![Diagram of cycle detection](image)

**Fig. 9** – detection of a cycle in order relations.

The existence of uncertain relations involves several cases of conflicts between relations: simple contradiction (with certain relations), contradictory uncertainty (with uncertain relations), invalidation (between uncertain and certain relations), or confirmation (case which is not a conflict, between uncertain and certain relations). If these cases are detected, they are specified on the datasheet.

Diagram manual modifications

As shown above, different diagram layouts are possible on the basis of the same set of stratigraphic relations. The shape of a diagram can be modified by taking account of different information (uncertain relations, chronological thresholds, phasing...). The introduction of these non stratigraphic elements reduces the number of possible diagram layouts with the same stratigraphic relations (BIBBY 2002). This appears for instance with the TPQ and TAQ thresholds mentioned above, which give the two possible extreme shapes of diagram compatible with the known data of absolute chronology. However, even if it includes these data, a diagram may have a great number of different layouts displaying different possible chronologies, because the site chronology (of which stratigraphy is only one part) is generally not completely known. Then, the
archaeologist has to choose, and to argue for, a chronological model compatible with stratigraphy and the other data. So, we think that the final layout of the diagram, which displays this model, comes under the choice of the archaeologist, for the best expression of both the stratigraphic sequence (which of course shouldn't be altered), and the chronological interpretation based on this sequence.

Practically, the drawing tools of Excel (or other software) may be used to modify the diagram obtained, and to add indications like frames showing units groups (features, structures…). For the future, we consider procedures of assistance to these modifications by user (for example by indicating when a modification is illogical).

Fig. 10 – Example of diagram obtained with Le Stratifiant, and then modified manually to show units groups (Qal’at Sem’an excavations, Syria, 2007).
Processing data from a database

*Le Stratifiant* may work with, or without, a local recording database or archaeological information system. If a database management system is employed to record stratigraphic observations and other field data (what is now frequent), *Le Stratifiant* can use the communication abilities of Excel to process data imported from the database, on condition that this database is able to generate and export two distinct tables: a table of units and a table of relations (pairs of units). That implies that the database must have a relational structure to manage units and stratigraphic relations.

Two main functions are at present allowed by this communication between *Le Stratifiant* and a recording database. The first one is generating diagrams with all the recorded units, or just with a selection of units; the second one is illustrating queries made in the database on the units, by coloring these units in the diagram (Fig. 11). Then, the diagram becomes a modifiable and dynamic document (like spatial representations in GIS).

![Fig. 11 - Detail from a diagram obtained with *Le Stratifiant*, with use of phasing, TPQ thresholds, and coloring functions. (Noyon, cathedral cloister – DESACHY 2008).](image)

Experimentation and development

*Le Stratifiant* is a work in progress: it is at present on trials on several archaeological excavations in France (Bibracte European archaeological centre in Burgundy, Noyon in Picardy, Chinon in the Loire valley) and Syria (Qal‘at Sem’an). These tests allow to find and solve bugs, and to define development prospects.
The application is contained in a PhD thesis (DESACHY 2008), in which the principles and the algorithm used by le Stratifiant are discussed. Both the thesis and the application are available (free download): http://tel.archives-ouvertes.fr/tel-00406241/en/
The current release is at present in French only, but an English version (or an attempt of it) is planned in the next months. Another evolution envisaged in a few months is a version for the free software Open Office.

Conclusion
Various choices are possible to design a tool for processing stratigraphic data. We think it is a good thing that several applications exist, like Stratify, Arched, Harris Matrix Composer, Le Stratifiant, and others, expressing these different choices of design and use. The choices materialized by Le Stratifiant - and thus its limits - are clear:
- choice of basic technical and software solutions, compatible with the ordinary not high level of computer equipment and practice which characterize the most part of the present archeological fieldwork (at least in France), and compatible with the rescue archaeology constraints.
- choice of basic and simple processing functions (and not advanced functions, like probabilistic chronological modeling, or 3D spatial processing...), in order to check and exploit more easily the recorded data, so that the archeologist may visualize quickly the whole results and logical consequences of his stratigraphic recording, including confrontation with absolute dating. This is a limited and practical aim, but we think it is a fundamental need (particularly in rescue archaeology conditions), to improve fieldwork management and results;
- choice to take in account the empirical nature of the archaeological approach, by including the distinction between "certain" an "uncertain" modes. "Empirical" does not mean "non scientific"; quite the reverse, we think this distinction certain / uncertain helps to keep the archaeological fieldwork rigorous and scientific;
- and, last but not least, choice to allow the user to have the last word. As it is said above, Le Stratifiant allows the archaeologist to visualize the consequences, and sometimes the contradictions, resulting from his observations. But the archeologist is responsible for these observations, and for their level (certain or uncertain); and finally, among the different possible layouts of diagram on the basis of the same stratigraphic structure, the final choice remains entirely under his responsibility. In short, the "philosophy" of Le stratifiant is: "for better or worse, the archeologist decides ; not the computer".

References

160 http://www.stratify.org
161 http://www.ads.tuwien.ac.at/arched/
162 http://www.harrismatrixcomposer.com


