An Automated System for Linear Feature Name Placement which Complies with Cartographic Quality Criteria

Mathieu Barrault - François Lecordix
Institut Géographique National - Laboratoire Cogit
2, Avenue Pasteur - BP 68 - 94160 Saint-Mandé France
barrault@cogit.ign.fr

ABSTRACT

Lettering is of prime importance for maps, but positioning is a time-consuming process which may account for up to 50% of the map-editing. This paper adresses the main parameters that lead positioning, and the problems entailed. Then an approach for automated road administrative name-placement is presented which takes into account these problems. We have implemented a research system which follows this approach. The results achieved so far indicate that the problem modeling we used allows to meet high quality requierements.

Keywords : Computational cartography, Optimization, automated name placement.

INTRODUCTION

Lettering is an important medium in cartography. Besides identifying geographic features, text parameters can convey classification (family), hierarchy (police size) and even feature location. But those information depend on the reliability of name text spatial allocation. Actually, this is an expensive task that may represent up to 50% of the map creation process [Yoeli, 72]. Although computer-assisted cartography developments have reduced processing time, only the automation of the label placement could improve on this cost sensibly.

The aim of this paper is both to analyze the main axes which constrain label placement, and to describe a program for automatic road administrative labels placement. The first part presents the three variables which constrain name placement and the problems they raise. After a quick paper review which clearly reveals the preponderance of the point-feature name placement, we'll focus on the problem of labeling linear features, and then bring in a system for automated road labels placement, developed at COGIT laboratory.

THE NAME PLACEMENT PROBLEM

A set of cartographic rules guides name placement [Imhof, 75] [Cuenin, 72]. Some change with the type of feature. Even if they do not provide a rigorous name placement theory, they reflect the three bases the label process leans on.

Three factors

A text attribute does not carry any explicit spatial location. But it cannot be positioned anywhere on a map because of three kinds of conditioners which guide the labeling. The first one is the feature the text designates. The name has to refer to it clearly (distance between the object and its name, text curvature...). Thus, most aesthetics rules depend on the type of feature (cf. figure 1).

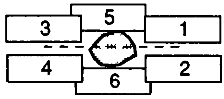

Figure 1 : A model for positional prioritization of point-features. [Cuenin, 72].

The second factor is the cartographic background i.e. every feature already placed and that cannot be moved to help the name to be well positioned. A name must not overlap point-features nor come to close to other significant features. Besides, name placement has to take into account the ambiguities the features of the cartographic background can produce (cf. figure 2).

The last factor is the set of other labels placed on the map. Two names must neither overlap nor be too close. What's more, they must not be evenly spread out nor densely clustered.

Problems of labeling.

Integrating those constraints raises problems we have to face in order to insure a suitable automated name placement under exacting cartographic constraints.

We can group criteria which don't rely on other names positions. They make up the *intrinsic quality* of a position. Cartographic requirements do not establish a strict hierarchy between them. They must be handled simultaneously. To achieve a significant intrinsic quality, we must find an exhaustive list of criteria and define a realistic measure to emulate the cartographer's assessment of each position a name can take. These criteria belong to the first two categories introduced above.

The other problem to face up is the label conflict, a NP-complete problem. What's more, a fine modeling has to manage ambiguities resulting from interactions between names and features. The last difficulty, but not the least, is to devise a system integrating all those without wronging any of them .

Progress in automated name placement

Several researchs have already been undertaken during the last 20 years. Most of them consider point feature names which represent the largest part of labeled features. Hirsch uses spatial search techniques to solve conflicts by computing an overlap vector from which the new position is derived [Hirsch, 82]. In the opposite way, a proposal is to produce a sorted set of possible positions for each feature and to pick up one effective position in each set so that none of them overlaps. This is a discrete approach which reduces combinatorial complexity. The effective position quest has different solutions. Sequential heuristics are proposed [Yoeli, 72] [Ahn & Freeman, 83] [Mower, 86] [Jones & Cook, 91] : each author has his own criteria to sort out features. Then, each feature is labeled : The best allowed position is used and conflicting positions are forbidden. If no position can be selected because none is available (cf. figure 3), there is a backtrack in the feature list to modify a previous selected position so as to free a possible position.

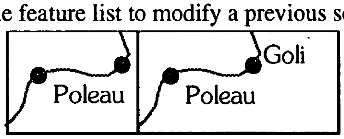

Figure 2 The link between Poleau and its feature depends on the legibility of the other name feature.

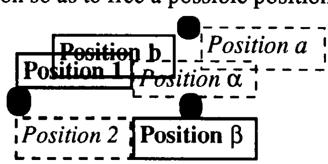

Figure 3 : Two conflictual labels (1,b). Solving this problem may induce another one and start a chain reaction even circular.

Cromley and then Zoraster proposed to consider name placement as a spatial allocation problem [Cromley, 86] [Zoraster, 87] that can be modeled as :

P_{ij} is a quality measure which takes its values in [0,1] where 0 denotes the best position and 1 the worst..

X_{ij}, is a boolean decision variable. $X_{ij}=1$ means the *j*th position is chosen for the *i*th point.

The distribution is modeled as the objective function : $$Min\sum_i\sum_j p_{ij}X_{ij}$$

with the constraint set :

Each point symbol demands a label $$\forall i, \sum_j X_{ij} \geq 1.$$

A set K of overlapping positions demands at most one position. $$\sum_{k \in K} X_k \leq 1.$$

Chirié and then Lecordix focus on the cartographic quality [Chirié, 92] [Lecordix & al, 94]. Destandau considers the road network [Destandau, 84]. She splits up long roads, needing to be labeled in more than one place, at specific nodes of the network she defines. Jones, who considers road networks too, focuses his works on the named features selection [Jones, 93].

This is a non-exhaustive literature review to introduce the main orientations.

LINEAR FEATURE NAME PLACEMENT PROBLEM

Firstly, the general guidelines listed in the previous section need to be instantiated for two classes of names. Nevertheless, point features name placement is the most investigated domain. In fact, it is the first issue which has been investigated at the IGN (see above), but our current research deals with linear features name placement. [Alinhac, 64] has defined such a set of cartographic rules for linear features.

-The name must be positioned along the linear feature it is labeling.
-It must be placed above the line.
-The label for a line feature should conform to the curvature of the line.
-Strong curvatures must be avoided. On the contrary, straight or almost straight parts must be preferred, preferably horizontal ones, the text being written from the left to the right.
-Small irregularities must be neglected.
-Vertical parts must be avoided. If not, in the left half of the map, the name must be placed on the left side of the line to read upward, and within the right part, it has to be placed on the right side to read downward.
-The name must not be separated from its linear feature by an other feature.
-The test must not cross the feature it is labeling.
-The name must not be spread out , but may be repeated at reasonable intervals along the feature.
-If needed, the name must be repeated along the feature. In this case, it must not be spread out or randomly repeated : The inter-distance should not vary dramatically.

Those rules show up new criteria we'll have to take into account : intrinsic quality criteria such as curvature, word spacing. Also, because of the long size a line can have, which can span across the whole map and because it almost belongs to a network (road network, river network....), the name should better be repeated along the feature. This is a good way to improve map legibility but the name placement process becomes more complex.

The main road network

To support this analysis, here is the description of our first automated name road network placement prototype, elaborated at COGIT laboratory. It is based on three guidelines :

- Road division in sections to label.
- Constitution of a finite set of possible positions (for each section) and computation of the intrinsic quality of those positions.
- Name placement by selecting among possible positions while taking into account intrinsic quality and label conflicts.

Repetition

Some roads are very long. This may cause a lack of legibility. The high density of neighboring features can affect its identification too. If the linear feature belongs to a network, nodes and other edges tend to obscure legibility. To retain the information all along the road, it must be repeated along the line. To process this repetition, a long road will be split into several sections according to specific criteria that must be labeled.

No section must be longer than an explicit size so as to bound the distance between two adjacent names on a road. Density is treated by restricting the number of nodes a section can hold. But this is not enough to comply with cartographic constraints. A crossroad can be embarrassing if sections are not well labeled. Some intersections, defined as a node where two edges of the road go through, are *dubious nodes* (cf. figure 4) because:

1. The two road's edges have different symbolization, inducing that it might be no more the same road. Otherwise :
2. The road crosses another road which is significantly thick. Again, the two edges may appear as belonging to two different roads.
3. Other roads with same legend cross our linear feature in an ambiguous way.

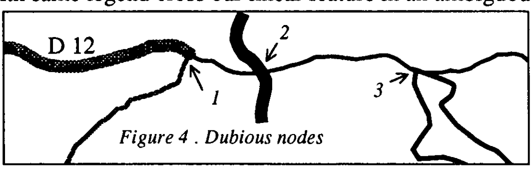

Figure 4 . Dubious nodes

We will call a *road unit*, a part of the linear feature which contains no dubious nodes except its ultimate nodes. Those road units are designed so as to have a reasonable length and to contain a limited set of nodes. In order to alleviate possible ambiguities induced by dubious nodes, every road unit has to be labeled. Labeling each road unit will clarify any ambiguity so that every road is easy to identified. But the map might be too dense. Some dubious crossroads, hence road units, can be neglected :

In fact, if both road units which precede and succeed a given road unit are labeled, then it may be unnecessary to label also this road unit (cf. figure 5).

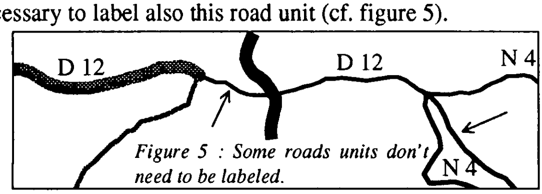

Figure 5 : Some roads units don't need to be labeled.

The main road network on small scale maps yields this kind of situation because of its structure (many tiny roads easily labeled, removing most ambiguities). To keep the message, road units are grouped so as to maximize nodes density and size without stepping over thresholds. These subsets are the *sections* of the roads. Each one of them has to be annotated.

Possible positions

The main road names are administrative labels (N 124, D 12). They are short text associated to artificial linear geographical features hence smooth linear features. Both characteristics allow us to restrict the prospecting of possible positions to straight parts and perform computations by using the bounding box of the label.

Prospecting possible positions provides a set of positions which can be subtituted to a conflicting position. An intrinsic quality [Lecordix & al, 94] must be defined and quantified for each possible position, at first to sort out the best possible positions. This intrinsic quality has to be normalized to permit the comparisons between possible positions of different features.

Restricted to straight parts, possible positions are tested by slipping the bounding box along the main axes of the linear feature. For each tried position, intrinsic quality's criteria are measured.

The main axes are computed with an algorithm issued from [Cromley, 92] based on the least mean square linear regression of the road sections.

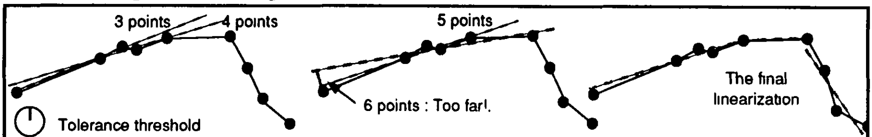

Figure 7 : A line simplification example.

Here are the intrinsic quality criteria for administrative labels.

- The name curvature : We have decided to restrict ourselves to straight parts so that we can use the bounding box of the label. The curvature is summed up to the angle the bounding box makes with the horizontal. This criterion is defined on $[-\pi/2, \pi/2]$.
- The axis : Most labels must be written above or below the line. Above is more suitable. Of course, if the name has to be placed on the center-line of the feature, this description does not apply.
- The local center : The more a position is centered between two crossroads, the finer the quality to be. To compute this, the bounding box is assimilated to its center of gravity. This point is projected on the line and its curvilinear abscissa is measured from one of the crossroads. This distance falls in [0, M/2] where M is the curvilinear abscissa of the other crossroad.
- The straight part : Because micro-inflections are neglected, we look for the straigthness of the labeled section. The less chaotic the section, the better the criterion. Assuming that the name and the line falls between S_{min} and S_{max} , this variable is defined in $[0, (l. (S_{max}-S_{min})]$ where l is the width of the bounding box.
- The background concealment : The label must avoid overlapping cartographic background. The more features overlap, the poorest the quality. Some features may even be forbidden. A position overlapping such a feature is called off. If the position does not conceal any prohibited feature, the background removal can be measured. The background map is locally rasterized and each pixel is weighted. The removal is computed by summing the cost of all pixels overlapped by the bounding box of the tried position. This measure is defined on [0, k.(N-1)] where k is the number of pixels in the bounding box and N is the highest possible pixel cost.

Each of those criteria partakes in the intrinsic quality. But, even if the cartographer can privilege some criteria among others, there is no strict theoretical hierarchy between them which could help to model the global intrinsic quality. They must be taken simultaneously and weighted. Another point is that extreme situations must be avoided. The valuation has to enhance these situations, we square each variable for this purpose.

So, for each possible position *j* of the section *i*, we can compute an intrinsic quality P_{ij}, depending on K criteria c_k as: $P_{ij} = \sum_{k=1,..,K} \alpha_k C_k(ij)^2$ with $\sum_k \alpha_k = 1$

$c_k(ij)$ being the estimated value of the *k*th criteria for the *j*th position the *i*th section, normalized on [0,1].

The higher the position quality, the smaller P_{ij}. 0 denotes the best position on a section.

Allocation

We have now a set of sections to label and for each, a sorted set of quantified possible positions. The size of the list must be a compromise between the result's quality and the processing time. We have to find the *effective position* of each section which will be labeled eventually. Actually, the placement of a label must take into account other names to guarantee a suitable legibility. In fact this problem must be considered at two levels. The local level at which we have to consider occurrences of the same name (repetition), and the global level at which all labels may interact.

Local allocation : labels for a specific road must be harmoniously scattered. The distance between any two successive effective positions must be roughly constant. Each label occurrence on the road must have a maximum intrinsic quality while joining in the regular allocation along the linear feature (cf. figure 8).

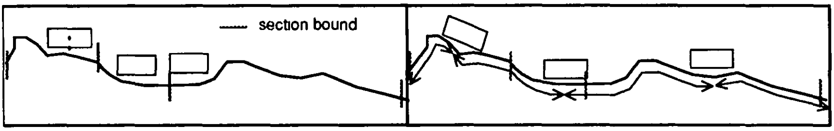

Figure 8 : Contribution of the local allocation.

Distance between two successive positions is calculated by the difference of the two curvilinear abscissae (cf. local center). Extremity labels must be near from the roads end and the others must be positioned so as to be regularly separated along the road while providing the best possible intrinsic quality. For a N-section road, we're looking for a set of effective positions $(P_1^e,..,P_N^e)$ which minimize :

$$\alpha \frac{\left(\left(d_{ac}(p_1)-\frac{n}{2}\right)^2+\sum_{i=1}^{k-1}\left(d_{ac}(p_{i+1})-d_{ac}(p_i)-n\right)^2+d_{ac}\left(p_k-\frac{n}{2}\right)^2\right)}{n^2(N+1)} + \beta \frac{\sum_i p_i^2}{N}$$

with n= Length T of the road / number of sections N.

$d_{ac}(p_i)$ the curvilinear abscissa of the *i*th section's tested position .

$\alpha+\beta=1$. A weighted sum to privilege one or the other quality criteria (regularity, intrinsic quality).

A simulated annealing followed by a discrete gradient descent gives a suitable solution.

Global allocation : Cartographic rules compel names neither to overlap each other nor to be too close. The proximity problem is assimilated to an overlap problem through expanding the bounding boxes. The position allocation must reject all overlapping expanded boxes.

Two different approaches have been used to tackle this NP-complete problem. The sequential approach is interesting for linear features because they have usually a lot of possible positions, and so that backtracking is rarely needed. But on the other hand intrinsic qualities worsen rapidly. Sequential methods are quick but they may yield a non optimal result. Moreover, the backtrack cost is higher with roads because changing an effective position will be passed on other label occurrences of the road, which may start up a chain reaction.

The holistic approach by Cromley & al [Cromley, 86] [Zoraster, 87], is theoretically adequate but the minimization is bound to be solved only through statistical heuristics (such as simulated annealing) because the objective function is neither derivable nor convex. In addition to this problem, taking into account the local allocation may complicate the function and lengthen the processing time needed for reaching a satisfactory solution.

So far, our method is "pseudo-sequential", since roads are processed one by one, while sections of the road are treated simultaneously. In our case, the sequential heuristic is sufficient because conflicts are rare and subsidiary positions handsome.
But in the long run, *global harmony* should also be considered by the name placement process. Currently, it is reduced to a boolean concealment. Actually, to improve name placement, it should involved the same kind of regularity we found in the local allocation, extended to the whole network.

Results
An experiment was performed on an extract of the French road map at 1:1,000,000 scale (Surroundings of Bordeaux - see figure 10). Local allocation is makes use of the simulated annealing method. Roads are processed one by one.
This extract is composed of 632 roads to be named, making 2125 edges.
The segmentation yields 1238 road units, grouped into 865 sections to be labeled.

It takes 2 minutes 30 to run the whole process with a DEC Alpha workstation. Over the whole French territory (26066 edges), it takes 1h40 to label 10187 sections.
The output of the extract was assessed by a cartographer from The Institut Géographique National and 88% of the positions were scored satisfactory. Further experiments will be carried on with other map sheets at other scales.

An automated mileage placement system derived from the method described above, has also been developed at the COGIT laboratory, based on this method [Marrot, 94]. Preprocessing includes the search of the best routes to be labeled. Once again, results are quite excellent with 85% of satisfactory mileages, within 3 minutes 10 of CPU time. Figure 10 illustrates place-names, road labels and mileages as positioned by the different systems developed at the laboratory (Note : Place-names are processed in 2 minutes, 80% of them being well-positioned).

CONCLUSION AND OUTLOOK

Even if lettering is a complex task, these encouraging results show that efficient automation becomes realistic. A common process for positioning different kinds of names shows up through these experimentations. However, the cartographic quality is highly dependent on demanding specifications that vary with the natures of the features.
Research will now focus on longer texts placed along linear features (such as river names...). Complexity increases with the amount of quality criteria and their measurement. Some texts are made up of several words. Besides repetition along the curve, they have to be positioned while taking into account each word's curvature and the space between two words... A modeling of the general problem similar to that described above

(which itself was deduced from [Lecordix & al, 94]) would be welcome. The ultimate aim is to find a global frame for name positioning, with specializations according to the types of names. This would provides with a common processing for names of different natures.

Another issue, besides local positioning, consists in addressing the relative positions between names, which is currently measured only in a boolean way (overlap or not). The limits of this approach are quickly revealed. This layout results into clusters of names which, even if not colliding nor overlapping, are nevertheless unsatisfactory. A finer measure is currently being studied so as to ensure global harmony between name-positions (for a given name category). The global harmony should eventually be extended to all kinds of features over the whole map so that all features may fit harmoniously and aesthetically into the map content.

References

[Alinhac, 64] G. Alinhac, Cartographie théorique et technique (theoratical and technical cartography), fascicule 1, chap. IV, I.G.N. Paris, 1964.

[Ahn, 84] J. Ahn, Automatic Map Name Placement system, Image Processing Laboratory Technical Report 063, Electrical, Computer, and Systems Engineer Department, Rensselaer Polytechnic Instiue, Troy, New York, 1984.

[Ahn & Freeman, 83] J. Ahn, H. Freeman, A program for automatic name placement, AUTO-CARTO VI, vol. 2, pp. 444-453, 1983.

[Barrault, 93] Placement automatique des toponymes sur le réseau routier au million, Rapport de stage de DEA, nov. 1993.

[Bonomi, 88] E. Bonomi, J.L. Lutton ,Le recuit simulé, Pour la Science, vol. 129, juillet 1988, pp.68-77, 1988.

[Chirié, 92] F. Chirié, Programme de positionnement automatique des noms de communes, Rapport de stage de DEA, sept. 1992.

[Cook et al, 91] A. Cook, C.B. Jones, J. McBride, Rule-Based control of automated name placement. Proceedings ICA'91 Bournemouth pp. 675-679, 1991.

[Cromley, 86] R.C. Cromley, A spatial allocation anaysis of the point annotation problem, Proceedings, Spatial Data Handling, pp. 39-49, 1986.

[Cromley, 92] R. Cromley, Principal axis line generalization, Computers & Geosciences col. 18, n. 8, p.1003 à 1011, 1992. [Cuenin,72] R. Cuenin, Cartographie générale, Tome 1, pp. 233-245, 1992.

[Destandau, 84] C. Destandau, Essai de positionnement automatique des numéros sur un réseau routier, Rapport de stage de fin d'études, IGN, 1984.

[Fairbain,93] D.J. Fairbain, On the nature of cartographic text, The Cartographic Journal, pp.104-111,Décembre 1993.

[Jones & Cook, 89] C.B. Jones, A.C. Cook, Rule-based name placement with Prolog, AUTO-CARTO 9, pp.231-240, 1989.

[Hirsch, 82] S.A. Hirsch, An algorithm for automatic name placement around point data, The American Cartographer, vol. 9, No. 1, pp.5-17, 1982.

[Imhof,75] E. Imhof, Positioning Names on Maps. The American Cartographer, vol. 2, pp.128-144, 1975.

[Jones, 93] C.B. Jones, Placenames, cartographic generalisation and deductive databases, NCGIA, october 93.

[Lecordix & al, 94] F. Lecordix, C. Plazanet, F. Chirié, J.P. Lagrange, T. Banel, Y. Cras, Automated name placement on map under high quality cartographic constraints, EGIS'94, vol.1, pp. 22-32, 1994.

[Marrot, 94] J.M. Marrot, Positionnement automatique des kilométrages, Rapport de stage de DESS, sept. 94.

[Mower, 86] J.E. Mower, Name placement of point features through constaint propagation, Proceedings Second International Spatial Data Handlings, pp. 65-73, 1986.

[Mower, 93] J.E. Mower, Automated Feature and Name Placement on Parallel Computers, Cartography and Geographic Information Systems, vol. 20, n. 2, pp. 69-82, 1993.

[Yoeli, 72] P. Yoeli, The logic of automatic map lettering, The cartographic journal, vol. 9, no. 2, pp. 99-108, 1972.

[Zoraster, 87] S. Zoraster, Practical experience with a map label placement program. Proceedings, Auto-Carto VIII. pp.701-707, 1987

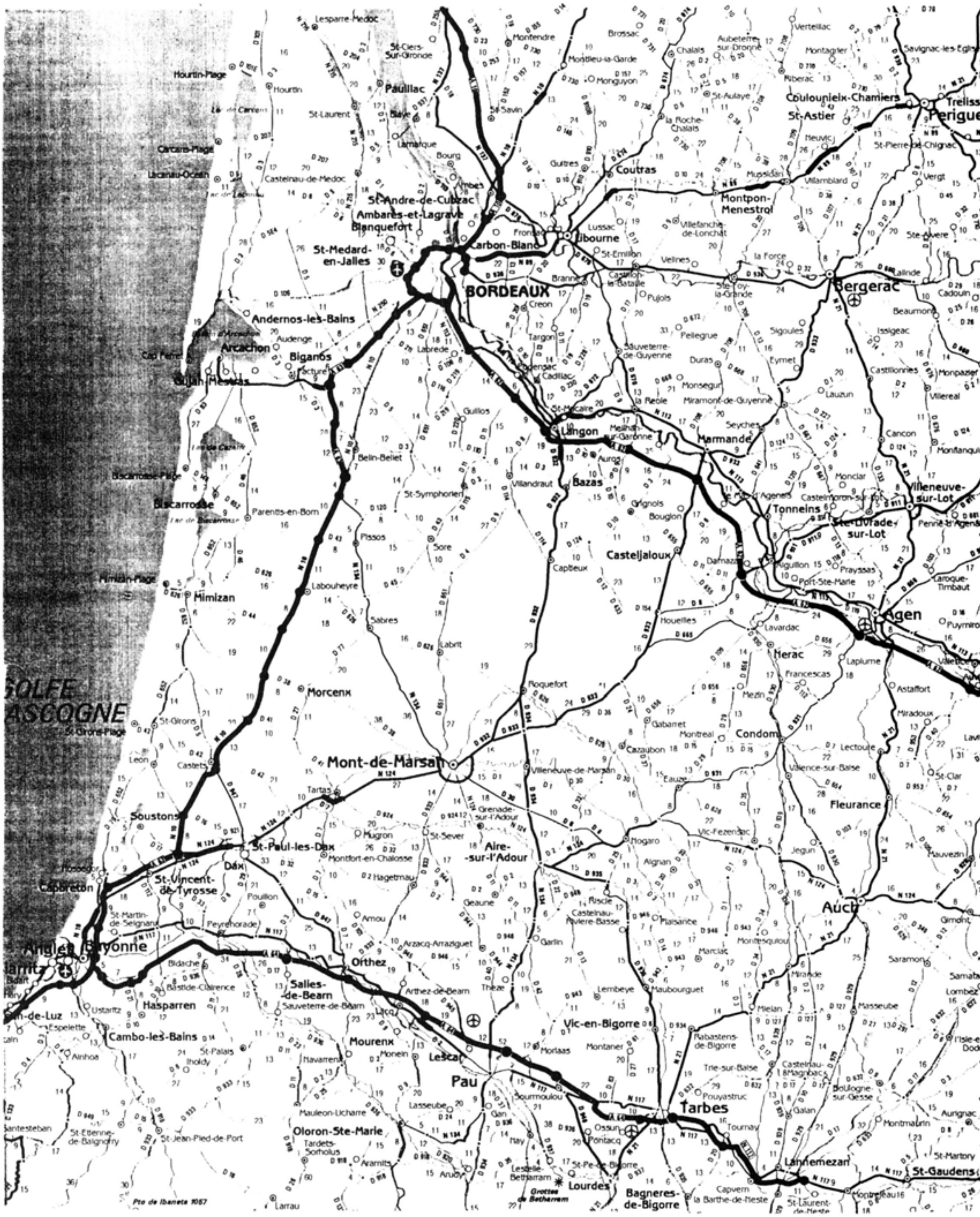

Figure 10 : Automated road label placement on an extract of the French road map at 1:1 000 000.