Lausanne Conference on Sequence Analyis (LaCOSA)
6th-8th June, 2012

François Buton, Claire Lemercier & Nicolas Mariot
CNRS, CEPEL-Center for the Political Study of Latin Europe, University of Montpellier-I;
CNRS, CSO-Center for the Sociology of Organizations, Sciences Po, Paris;
CNRS, CURAPP-University Research Center on Public Action and Policy,
University of Picardie Jules Verne, Amiens

A Contextual Analysis of Electoral Participation Sequences

This presentation is mostly based on a paper that is soon to be published in *Electoral Studies*; the full paper is already available online at:

http://www.sciencedirect.com/science/article/pii/S0261379411001429.

However, the presentation will put more emphasis on the use of sequence analysis in this research and will include additional, still unpublished results. We study sequences of electoral turnout, coded from signature lists. Our sequences are hence both very simple and very complicated.

On the one hand, we only have information on turnout, not on the party or candidate that was chosen. For each election, there are only three possible states: participation, abstention, and non-registration. As our dataset is based on the signature lists of one polling station, observed between 1982 and 2007, that had 500 to 800 registered voters for each election, the "non-registered" state could occur when one person who was registered in this station for at least one election in this period was too young to vote (or dead) or registered in another station. Signature lists do not allow to observe many covariates that would likely help to explain turnout; they especially do not include any direct information on the voters' occupation, wealth or diploma. However, we were able to gather information or to construct reasonable proxies for several covariates: we know each voter's sex, date of birth, exact address (and we know the social characteristics generally associated with each neighborhood) and date of first registration in the polling station when it happened after 1982 (which is an indicator of the fact of having moved to the place that we observe). As our polling station is located in a *ville nouvelle*, or « new town », that has transformed from a rural village to a part of the Paris suburbs mostly inhabited by the higher and middle class, this proxy of migration is particularly interesting. Being based on this particular case study, our substantive results arguably

cannot all be generalized, but we provide a methodology based on an innovative use of signature lists and sequence analysis that will hopefully allow comparisons and generalizations in the near future.

On the other hand, our dataset is indeed uncustomarily rich as regards two important dimensions. First, we observe no less than 29,756 direct traces of acts of turnout or abstention: while surveys on turnout often prove unreliable, we have proof that the registered voters actually did or not not participate, and we follow them through 44 ballots, corresponding to 27 different local or national elections. This time span is unparalleled in previous studies: a few researchers had already pointed out that constant participationists and abstentionists were scarce, but they always observed less than 10 ballots. On the contrary, we have long-term and fine-grain data that also allow us to differentiate behaviors according to the type of election.

Second, thanks to the fact that collective housing is rare in our town, we were able to trace the existence of what we call "electorate households" in our data. "Electorate households" are defined as groups of registered voters living together and generally akin to each other; only registered voters are taken into account in this definition, so that e.g. the mother of three children under 18, married to a foreign citizen, would be considered as isolated (member of a household of one) in our definition. Our study takes advantage of our detailed data on turnout and electorate households to show that turnout is highly correlated inside electorate households, so that a/ the behavior of other members of the electorate household and b/ the status of the voter in an electorate household are very good predictors of individual turnout. Studies of turnout only based on individual covariates are therefore likely to miss their point.

In the first part of our *Electoral Studies* paper, we first use ANOVA and pseudo-ANOVA to show that electorate households are much more homogeneous than chance would allow as regards several dimensions of turnout behavior. In the second part, we discuss the effect of the status in the electorate household on participation: this status is defined as being either the parent of a voter, part of a couple of voters (not a parent), in another situation, the child of a voter (not part of a couple), or isolated. Overall turnout rates are decreasing from parents to isolated voters, implying a correlation between social integration among registered voters and turnout. This is still true in a multilevel model of overall turnout: this model confirms the weight or intra-household correlations and, all other things being equal, the higher participation of parents of voters and the lower participation of isolated voters. In addition, our model shows differences in participation related both to the date of birth and the date of registration, that are likely to capture a social differentiation between newcomers in the new town and former inhabitants.

This paper partly relies on techniques that have nothing to do with sequence analysis per se: the

explained variable in the final model is a turnout rate calculated over all ballots for those who were registered for at least four ballots, and the covariates are static. However, sequence analysis played an important role in our exploratory research for this paper:

- graphs provided by TraMineR, although we could otherwise have approximated them with Excel or R, offered a quick and easy way to navigate our data and to think about dimensions of homogeneity in households – and in fact to convince ourselves that the very high homogeneity rates revealed by calculations were not a mere artifact. However, it would be useful to discuss the effect of such sequence graphs on readers: as they represent each ballot by the same bar, they tend to obliterate the representation of calendar time;
- general ideas from the sequences literature allowed us to devise three different indexes describing sequences and to test the homogeneity of households for each of the three indexes. In addition to the overall turnout rate, the classical index in electoral sociology (and to overall turnout rates for local vs. national elections, that we are currently investigating), we wanted to use other indexes taking the precise temporality of behavior into account. We first used a "change of behavior index", differentiating those who alternated between participation and abstention at each ballot or election from those who changed one time and for good. This is a useful refinement as compared to rough descriptions of "temporary abstentionists" previously used in the literature; it could easily have been calculated without TraMineR, but it is part of the realm of sequence ideas. We were able to show that this index was highly correlated inside households, and that while the younger generations exhibited higher change indexes, being the parent of a voter also significantly decreased the change index, all other things being equal;
- our third index measures the similarity between couples of sequences in terms of exact timing of turnout. This is allowed by the use of optimal matching, with a specific set of costs. What we want to measure here is (especially at the scale of the household) the tendency not only to generally participate, but to participate or to abstain for the exact same ballots. A simple count of exact matches between sequences would thus have served our purpose if all voters had been registered for the whole period, but this was not true. We hence used an optimal matching distance with the highest possible indel costs (as we did not want to have similar sequences occurring at different dates considered as close), low substitution costs between participation and non-registration and between abstention and non-registration, and high substitution costs between participation and abstention. Visual inspection confirmed that this measured what we wanted to measure. The pseudo-ANOVA calculation provided by TraMineR then allowed us to prove that behaviors inside electorate households were also much more correlated in this respect than chance would allow; we controlled the robustness of this result by excluding the smaller households from

the calculation (what we measure is not only homogeneity inside couples) and by comparing homogeneity inside households and inside addresses (which is significant but lower).

In the LaCOSA conference, we hope to discuss the choices and results presented above with specialists of sequence analysis. In addition, we will present and discuss two types of additional results:

– a tree diagram splitting the exact voting sequences (using our optimal matching distance) according to several covariates, thus extending our multivariate discussion of factors influencing turnout to its precise timing. In our *Electoral Studies* paper, we only gave a detailed discussion of factors influencing overall turnout and hinted at influences on change of behavior, but we did not publish the tree diagram – it would have been too much to explain in a paper that already introduced sequence analysis to a new audience. The graph however gives very interesting results, as it shows precise interactions between the date of birth, date of registration and status in the household – which were also the three main predictors of overall turnout and change in behavior. Depending on generations, it is parents of voters, children and parents or isolated voters that have peculiar behaviors. This result however should probably be taken with caution, as it describes registration as much as turnout and as it is not possible, to our knowledge, to include a household variable in a multilevel model, while our other results show that correlation inside households is very important;

– a study of the transmission of electoral behavior and especially of the impact of becoming the parent of a voter on participation. A preliminary study of the overall turnout rate showed that the rate for a child of voters could quite accurately be predicted by the average rate of his or her parents. However, this is partly a *post hoc* prediction, as these rates were in fact measured for contemporary periods. In addition, our finding that parents of voters tend to participate more than couples of voters is also based on a static coding of voters as parents, even if their children only became voters at some point during the observed period. We thus want to investigate more precisely what happens when the composition of electorate households change, which mainly occurs when childrens of voters turn 18. Are the rate of turnout and change of behavior index of children becoming voters actually predicted by those of their parents *before* they became voters? And do the parents participate more frequently after their children have become voters themselves? This would point to interesting mechanisms, helping to open the black box of correlation inside households. At this stage, we will perform exploratory analyses of this dimension of our data, mainly using simple calculations and visualizations at the scale of the household and the dataset, and borrowing the idea of aligning sequences according to an external date (here, that of the registration of children as voters) to Simon Paye and Denis Colombi.