

## APPENDIX G:

### SCATTER PLOTS OF FINAL TEST RESULTS

This appendix contains selected scatter plots of much of the data found in appendix F. The correspondence between the data points in appendix G and the tables in appendix F is as follows:

Appendix G Data Point Label	Appendix F Table Number
ADS	F1
BBN	F2
BBN-O	F3
GE	F4
GIE	F5
HU	F6
ITP	F7
LSI	F8
MDC	F9
NYU	F10
NYU-O1	F11
NYU-O2	F12
PRC	F13
SRI	F14
SYN	F15
UNI	F16
UMA	F17
UMA-O	F18
UNL	F19
UNL-O1	F20
UNL-O2	F21

## SECTION 1. OVERALL RESULTS

The plots in this section are taken from the four summary rows that appear at the bottom of the score reports. Discussion of some of these plots is contained in the paper by Sundheim in Part I of this proceedings.

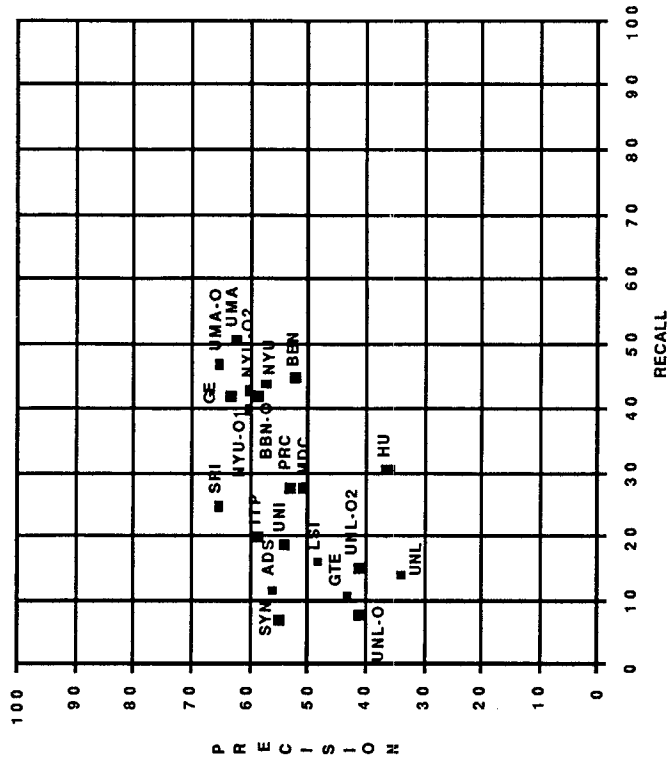


Figure G1. Overall Recall vs Precision: Matched/Missing Row. These scores include moderate penalties for missing and spurious data.

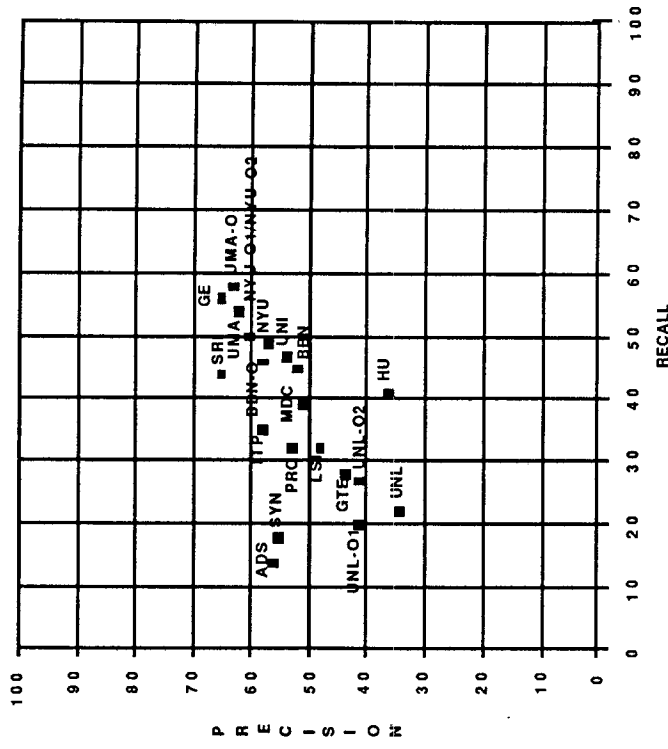


Figure G2. Overall Recall vs Precision: Matched Only Row. These scores include only mild penalties for missing and spurious data.

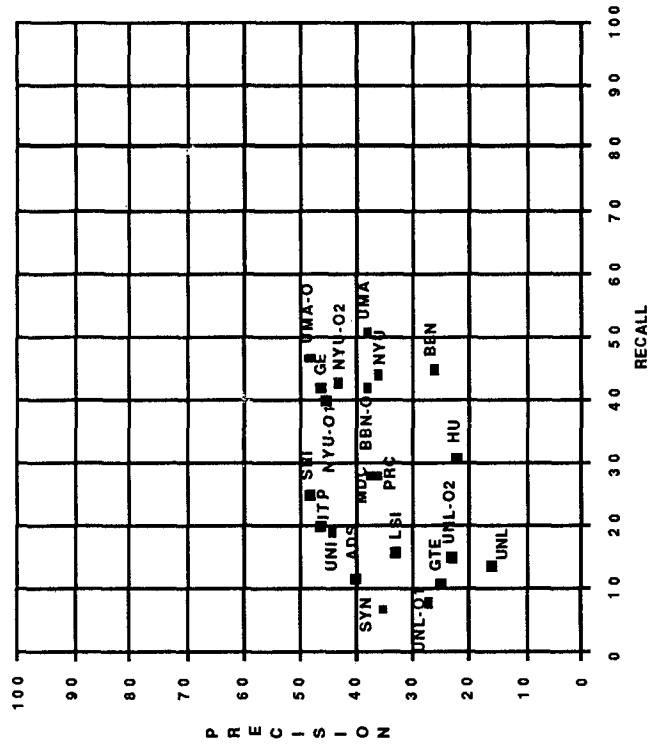


Figure G3. Overall Recall vs Precision: All Templates Row. These scores include severe penalties for missing and spurious data.

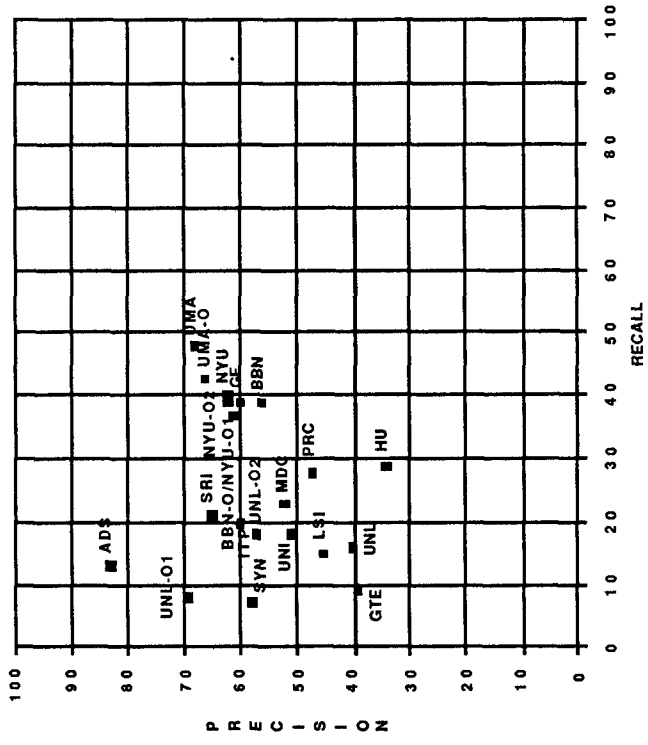


Figure G4. Overall Recall vs Precision for the Set Fills Row. This "overall" method of computing the scores is limited to those slots requiring set fills (fills from a set of predefined alternatives).

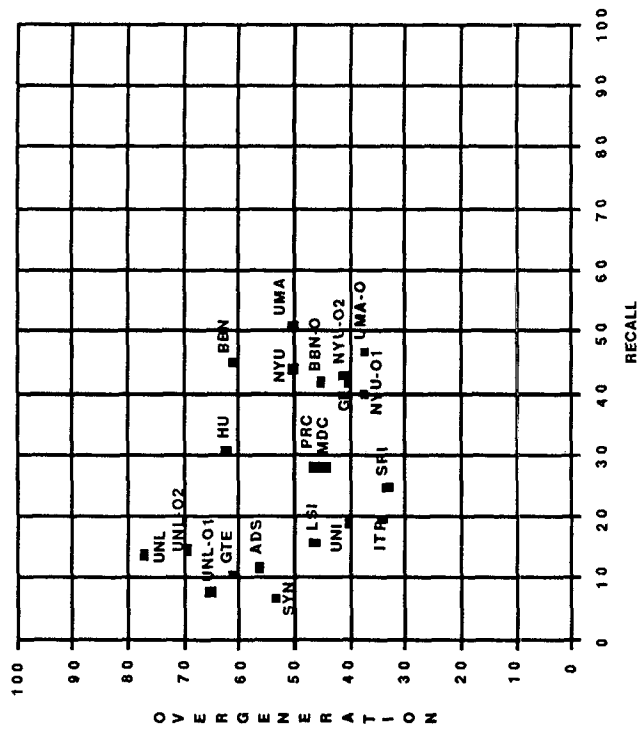


Figure G6. Overall Recall vs Overgeneration: All Templates Row. This plot shows how much spurious data generation is taking place, given the severe All Templates method of assessing penalties.

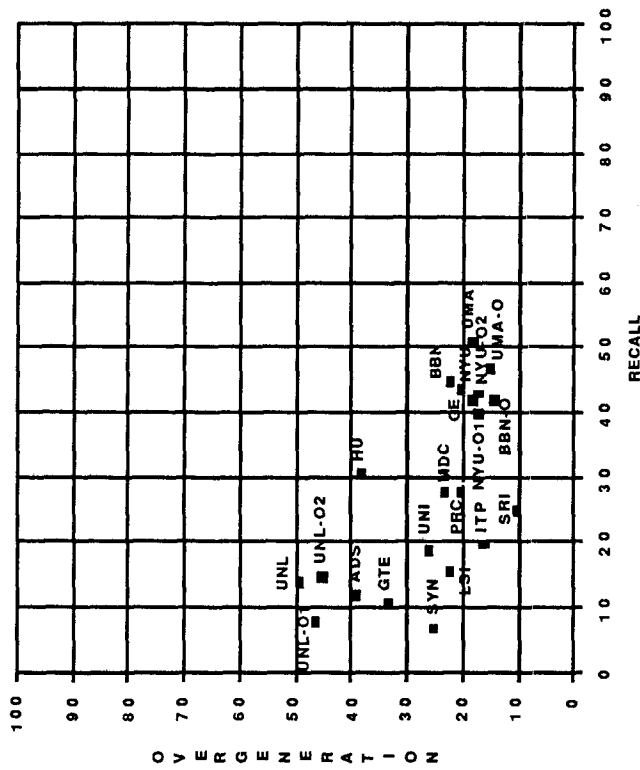


Figure G5. Overall Recall vs Overgeneration: Matched/Missing Row. This plot shows how much spurious data generation is taking place, given the Matched/Missing method of assessing penalties for missing and spurious data. (The values on the Overgeneration axis would be the same under the Matched Only method.)

## SECTION 2. SLOT-BY-SLOT RESULTS

The plots in this section are taken from the top part of the score reports, which contains the results for each slot in the MUC-3 template. Discussion of a few of these plots is contained in the paper by Sundheim in Part I of this proceedings. Some systems did not attempt to fill all the slots in the template; thus, not all systems appear in all the scatter plots.

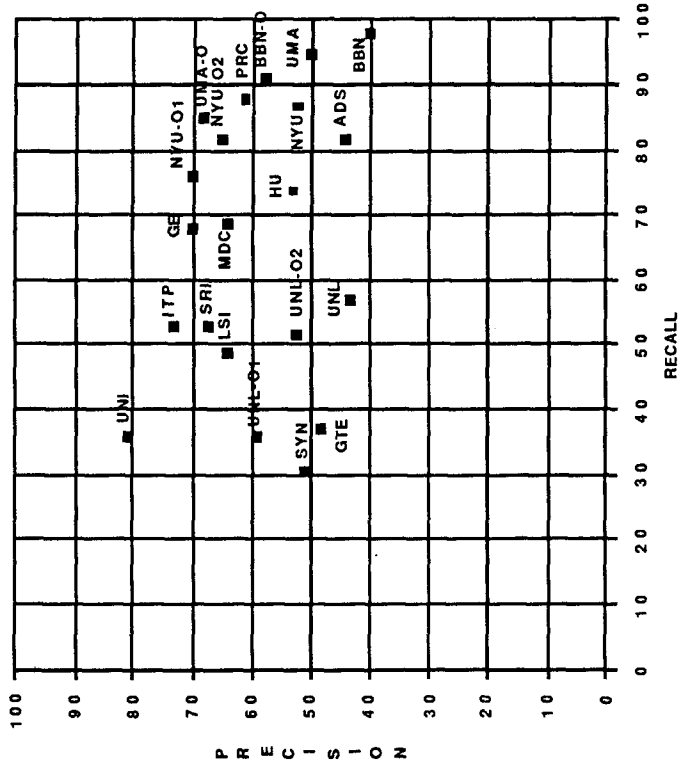


Figure G7. Slot 1 (template-id) Recall vs Precision

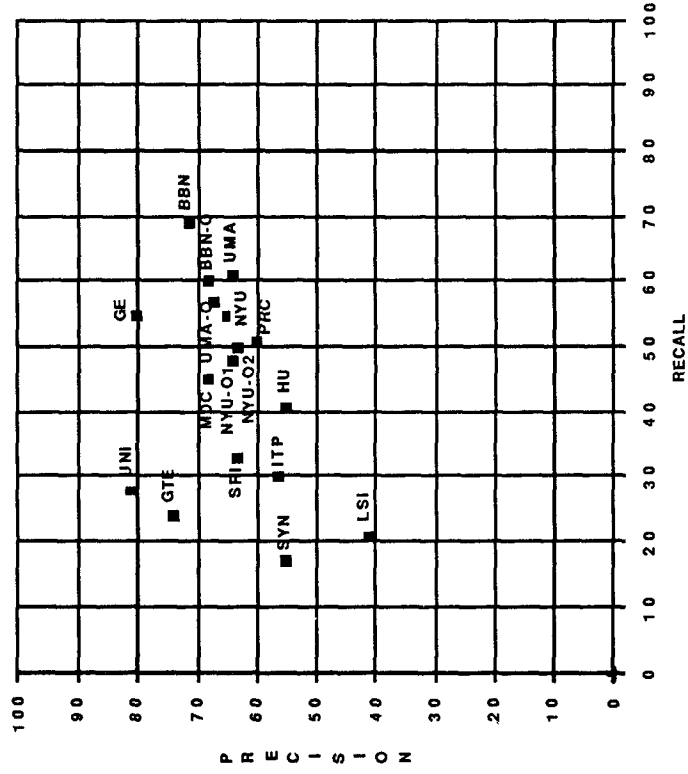


Figure G8. Slot 2 (incident-date) Recall vs Precision

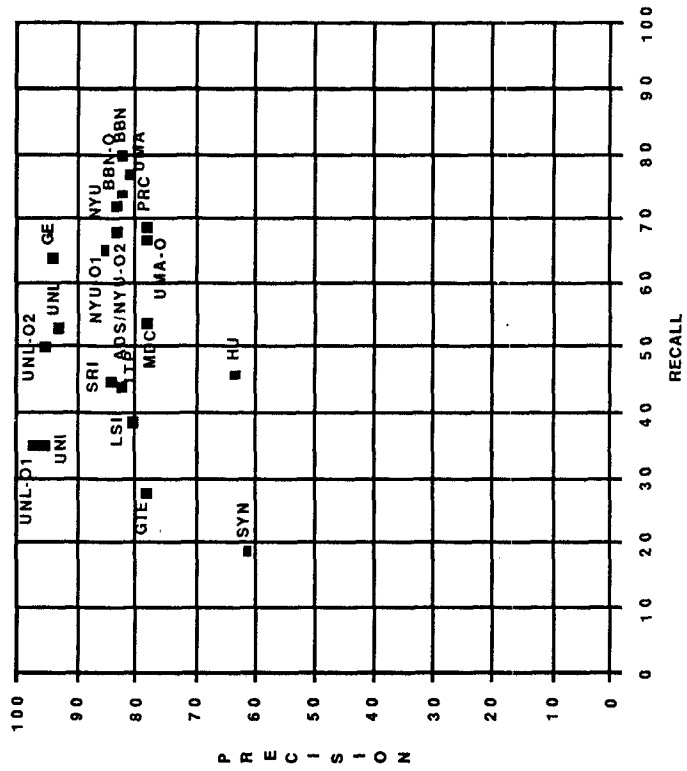


Figure G9. Slot 3 (incident-type) Recall vs Precision

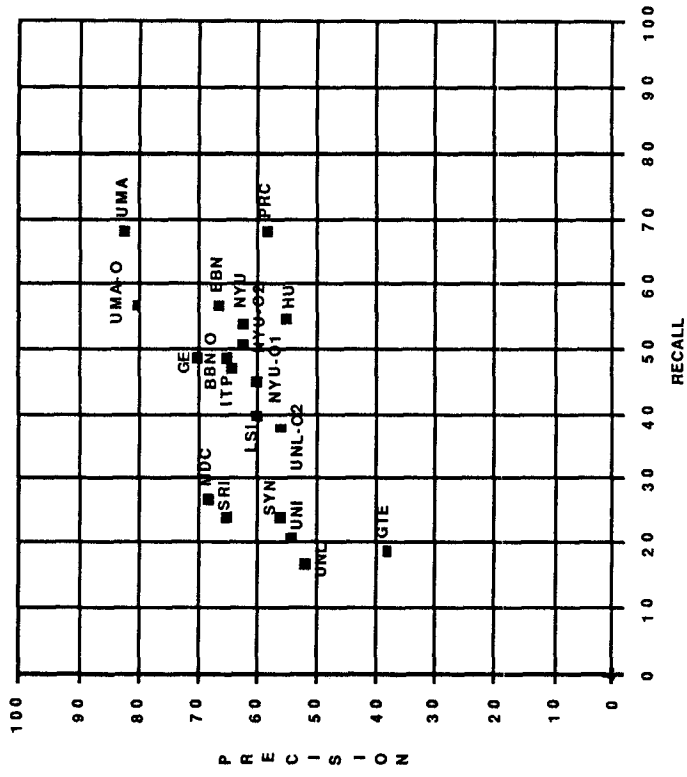


Figure G10. Slot 4 (category) Recall vs Precision

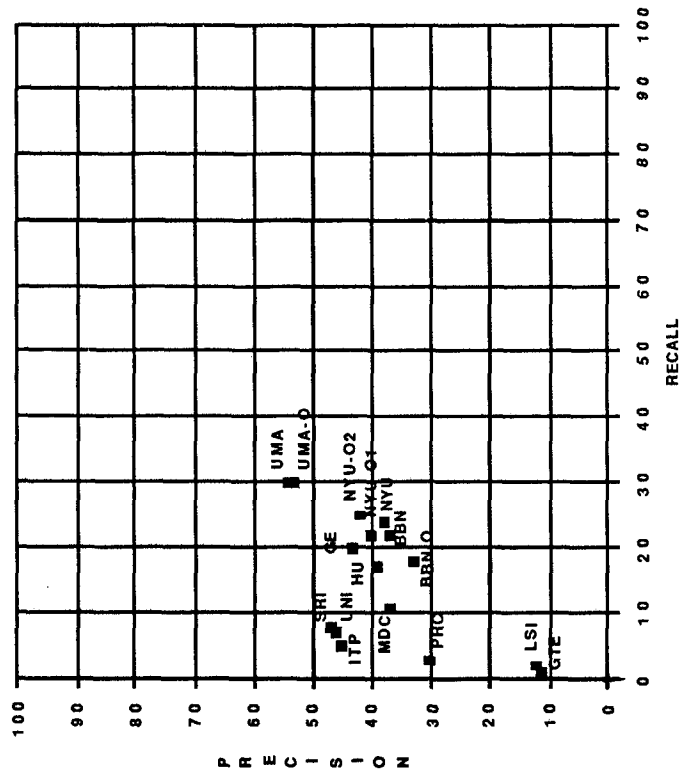


Figure G11. Slot 5 (indiv-perps) Recall vs Precision

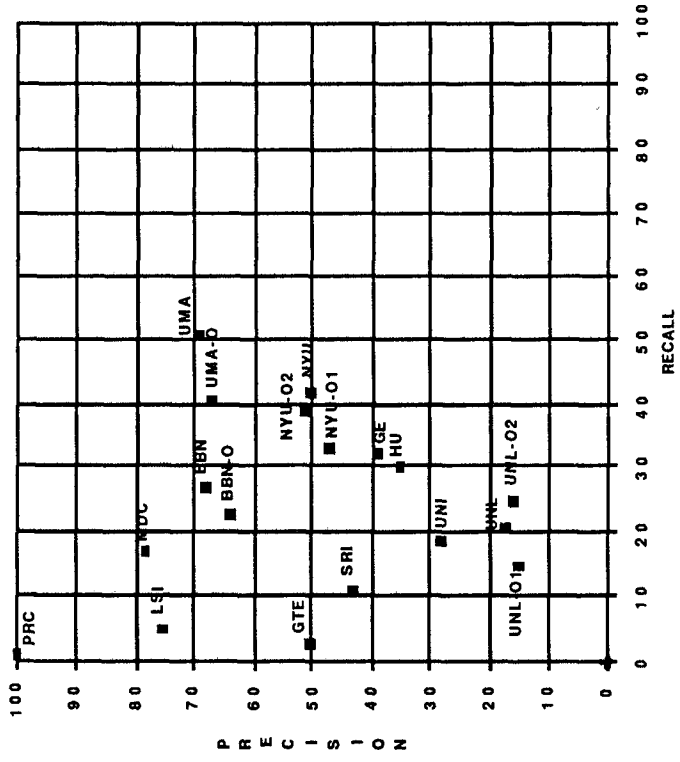


Figure G12. Slot 6 (org-perps) Recall vs Precision





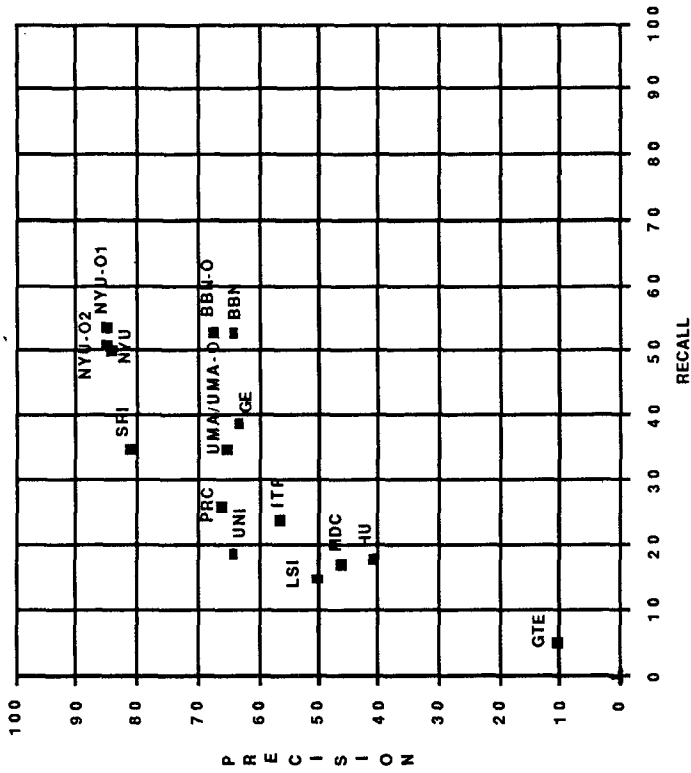


Figure G15. Slot 9 (phys-target-num) Recall vs Precision

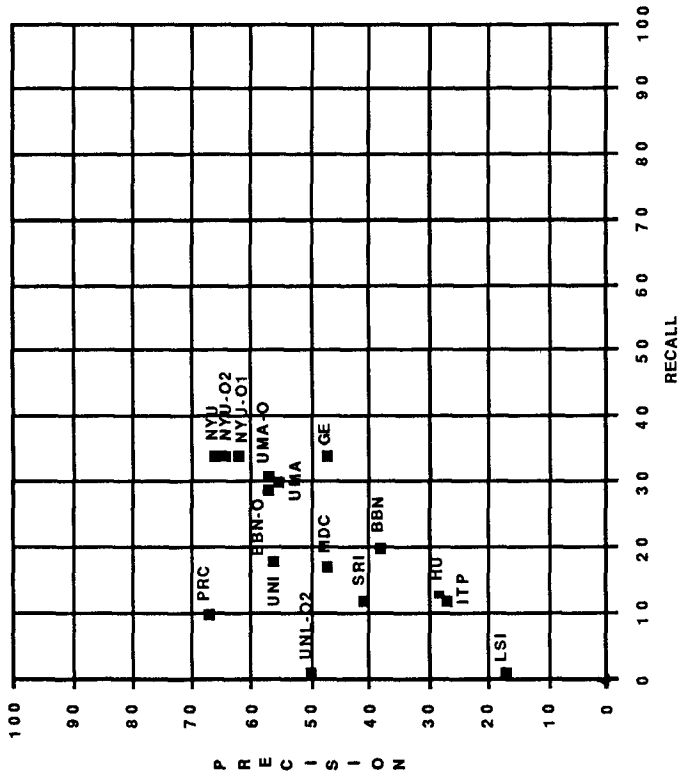


Figure G16. Slot 10 (phys-target-types) Recall vs Precision

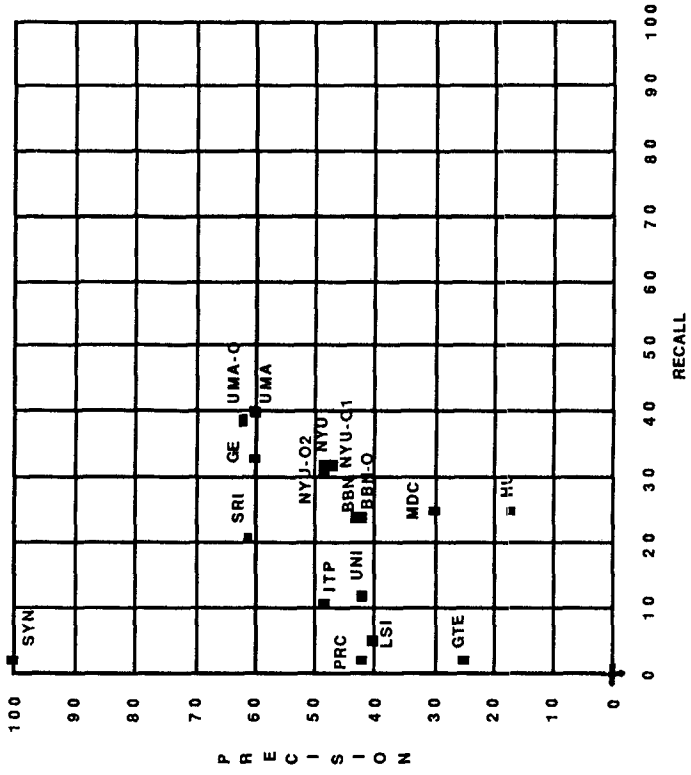


Figure G17. Slot 11 (human-target-ids) Recall vs Precision

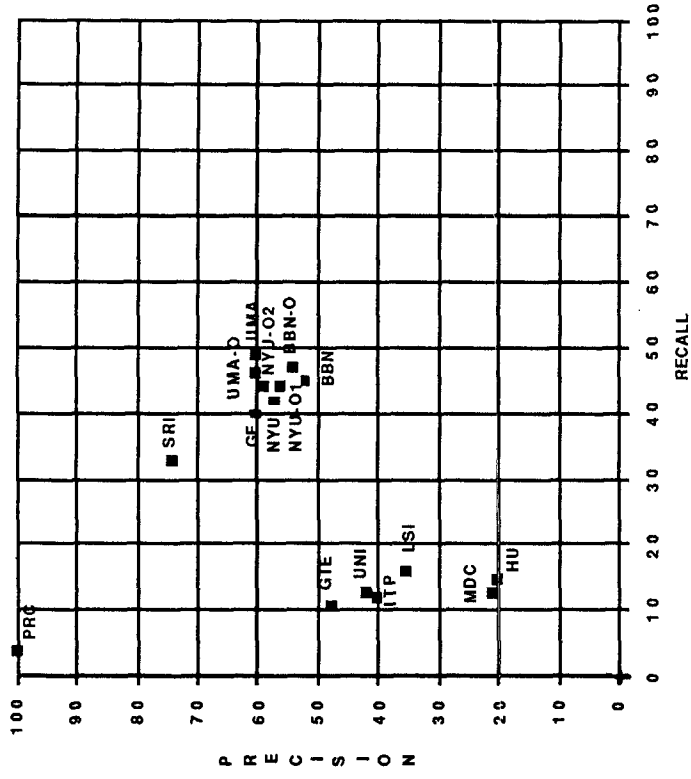


Figure G18. Slot 12 (human-target-num) Recall vs Precision

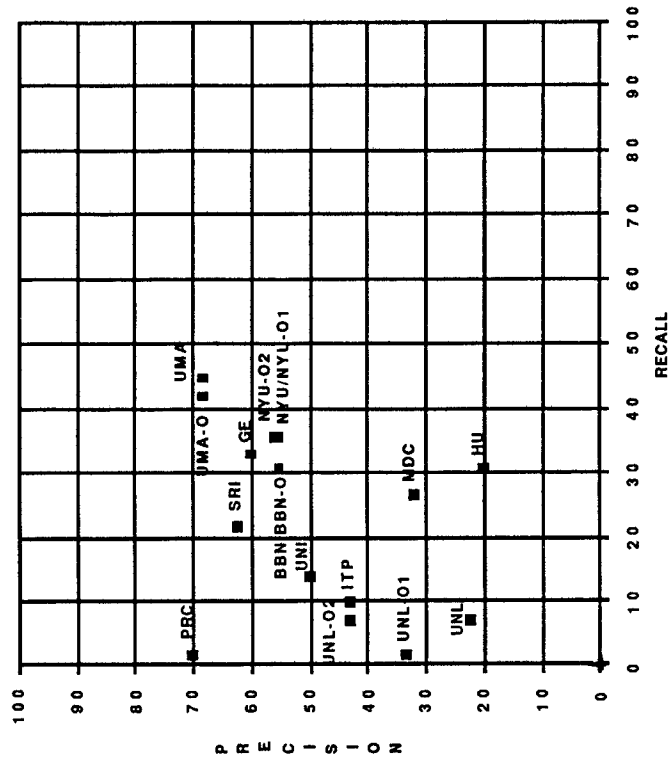


Figure G19. Slot 13 (human-target-types) Recall vs Precision

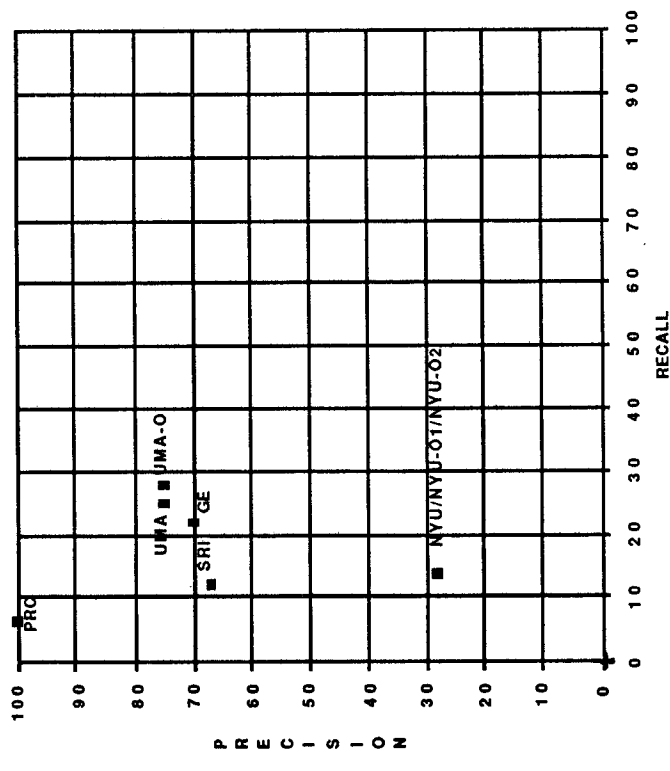


Figure G20. Slot 14 (target-nationality) Recall vs Precision

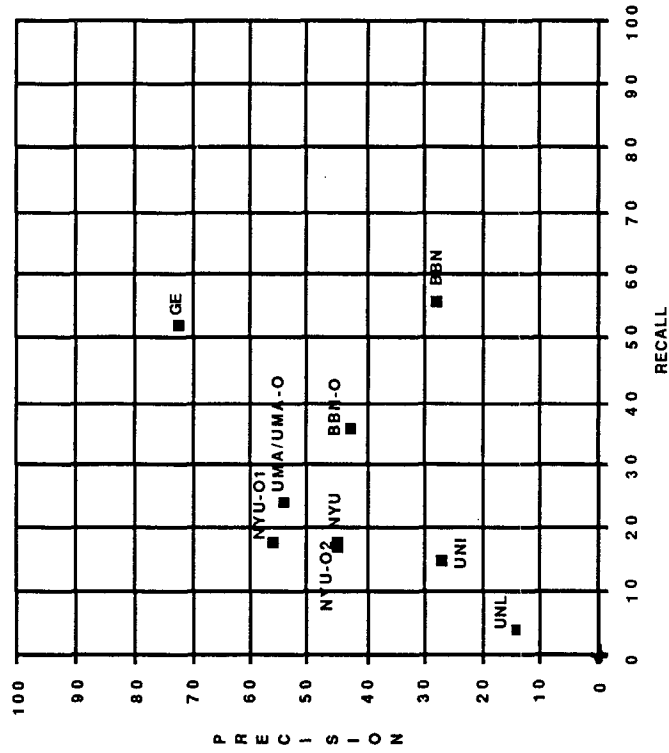


Figure G21. Slot 15 (instrument-types) Recall vs Precision

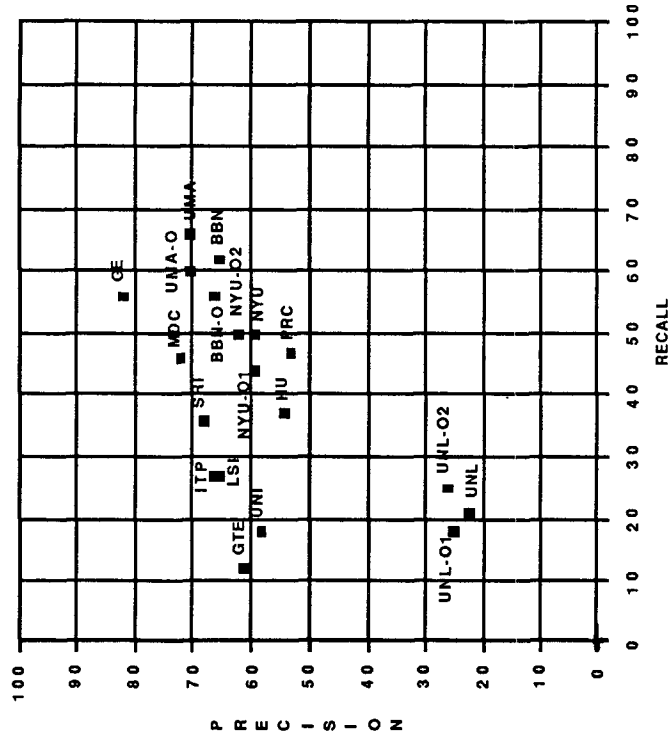


Figure G22. Slot 16 (incident-location) Recall vs Precision

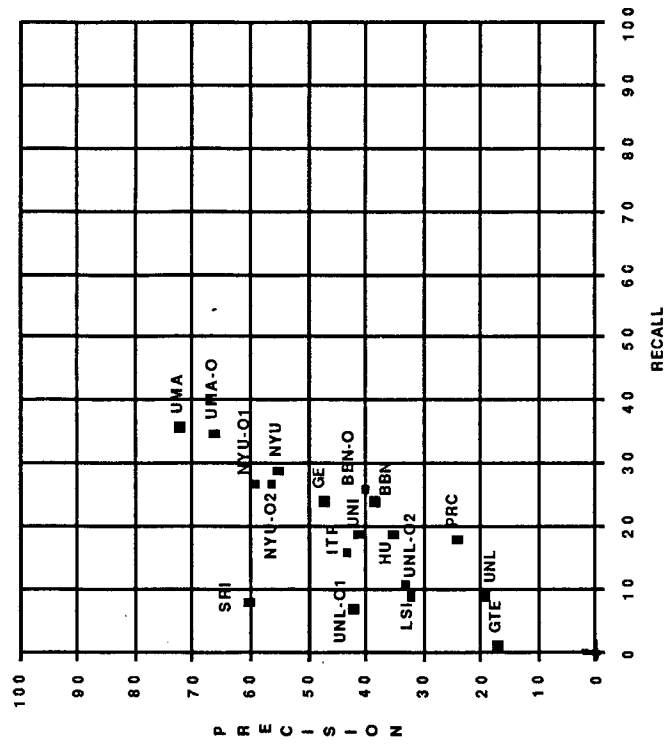


Figure G23. Slot 17 (phys-effects) Recall vs Precision

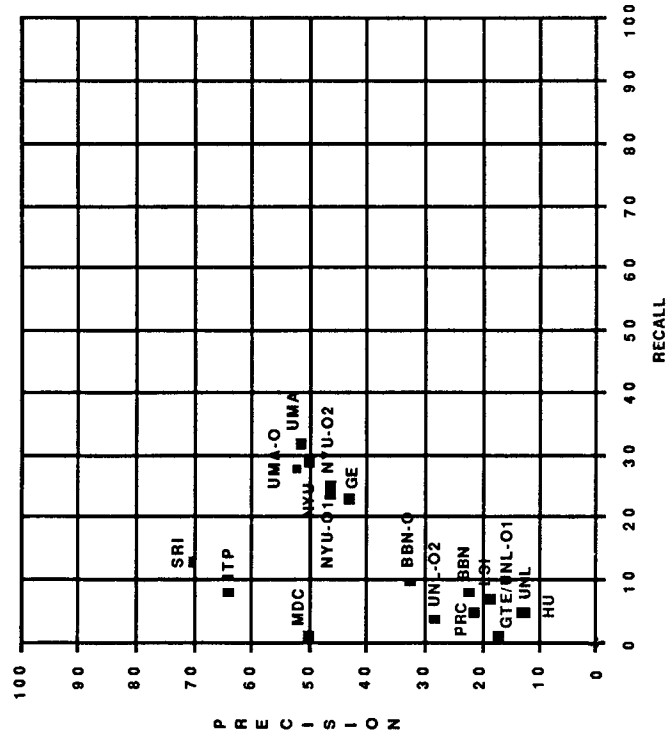


Figure G24. Slot 18 (human-effects) Recall vs Precision