

| Masking Rates |      |      | Dev Set Results |           |               |
|---------------|------|------|-----------------|-----------|---------------|
| MASK          | SAME | RND  | MNLI            | NER       |               |
|               |      |      | Fine-tune       | Fine-tune | Feature-based |
| 80%           | 10%  | 10%  | 84.2            | 95.4      | 94.9          |
| 100%          | 0%   | 0%   | 84.3            | 94.9      | 94.0          |
| 80%           | 0%   | 20%  | 84.1            | 95.2      | 94.6          |
| 80%           | 20%  | 0%   | 84.4            | 95.2      | 94.7          |
| 0%            | 20%  | 80%  | 83.7            | 94.8      | 94.6          |
| 0%            | 0%   | 100% | 83.6            | 94.9      | 94.6          |

Table 8: Ablation over different masking strategies.