



The idea is to first split each of the three files into three parts (shown with shading on the left) which we will call  $X$ -train,  $X$ -dev,  $X$ -test, where  $X \in \{\text{male, female, mixed}\}$ . Then we'll create three different evaluation scenarios:

1. Target=male; shown top. You will train using the union of subsample(male-train), female-train, mixed-train. You will use for development subsample(male-dev), and for testing male-test. The function subsample returns a 100-instance subset of the dataset.
2. Target=female; shown bottom. You will train using the union of male-train, subsample(female-train), mixed-train. You will use for development subsample(female-dev), and for testing female-test.
3. Target=mixed; not shown. You will train using the union of male-train, female-train, subsample(mixed-train). You will use for development subsample(mixed-dev), and for testing mixed-test.

Perhaps I shouldn't have called it cross validation, as although it's related, it's a little different from standard cross-validation. We're basically forming a data-improverished scenario for each target domain for each scenario, and letting each of the domains serve as the target.