1 Reporting a binomial test

1. **R expression(s)**

```
> x <- 501
> n <- (501 + 1859)
> p <- 0.5
> binom.test(x, n, p)
            Exact binomial test
    data:  x and n
    number of successes = 501, number of trials = 2360, p-value < 2.2e-16
    alternative hypothesis: true probability of success is not equal to 0.5
    95 percent confidence interval:
     0.1959431 0.2293504
    sample estimates:
    probability of success
            0.2122881
```

2. **Test Report**

   This experiment looks at a corpus of 501 prepositional datives constructions and 1,859 double object constructions and has a null hypothesis assumption that the two constructions are equally likely to occur. Using a binomial test, we can reject the null hypothesis because the probability of success (x/n) is 0.2122881, which does not fall within the 95% CI of 0.1959431 and 0.2293504. Furthermore, the p-value is < 2.2e-16, which is significantly smaller than α = 0.05.

2 McNemar's Test

1. **R commands**

```
> ptbfile <- read.table(file = 'C:/Users/wyan3/OneDrive/CUNY GC/Statistics/PTB.tsv', sep = '\t',
header = TRUE)
> Stanford.correct <- ptbfile$gold.tag == ptbfile$Stanford.tag
> NLP4J.correct <- ptbfile$gold.tag == ptbfile$NLP4J.tag
> x1 <- sum(Stanford.correct & ! NLP4J.correct)
> x2 <- sum(NLP4J.correct & ! Stanford.correct)
> x1
    [1] 943
> x2
    [1] 1016
> x <- x1
> n <- (x1 + x2)
> p <- 0.5
> binom.test(x, n, p)
            Exact binomial test
    data:  x and n
    number of successes = 943, number of trials = 1959, p-value = 0.1038
    alternative hypothesis: true probability of success is not equal to 0.5
    95 percent confidence interval:
     0.459029 0.503763
```

```
sample estimates:
probability of success
         0.481368
```

2. **Number of wins for Stanford tagger over NLP4J tagger**
   The Stanford tagger has 943 wins over the NLP4J tagger.
3. **Number of wins for NLP4J tagger over Stanford tagger**
   The NLP4J tagger has 1,016 wins over the Stanford tagger.
4. **McNemar test results; is one significantly better than the other at $\alpha = .05$? If so, which one?**
   From the McNemar test results, it can be determined that neither tagger, Stanford or NLP4J, is significantly better than the other. We cannot reject H_0 because the p-value of 0.1038 is greater than $\alpha = 0.05$, and the probability of success ($x/n = 0.481368$) falls within the 95% CI of 0.459029 and 0.503763.