**Supplementary Materials: Statistics**

Supplementary Table 1

*Wilcoxon test statistics for comparisons of clause ordering (main>subordinate versus subordinate> main) in habitual and ongoing contexts*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | CMT-Listen  Accuracy | CMT-Read  Accuracy | SPR critical word | SPR whole sentence |
| Habitual |  |  |  |  |  |
| Pretest | Z  p (*d*) | -.265  .791 (.08) | -.212  .832 (.06) | -.169,  .866 (.04) | -.789  .430 (.22) |
| Posttest | Z  p (*d*) | -1.202  .229 (.35) | -.005  .996 (.00) | -.845  .398 (.24) | -.976  .309 (.28) |
| Delayed Posttest | Z  p (*d*) | -1.269  .204 (.37) | -.082  .935 (.02) | -.941  347 (.26) | -.909  .364 (.26) |
| Ongoing |  |  |  |  |  |
| Pretest | Z  p (*d*) | -1.165  .244 (.32) | -.039  .969 (.02) | -.203  .839 (.06) | -.967  .333 (.28) |
| Posttest | Z  p (*d*) | -1.289  .197 (.37) | -.082  .935 (.02) | -.941  .347 (.26) | -1.29  .194 (.37) |
| Delayed Posttest | Z  p (*d*) | -1.345  .179 (.32) | -.043  .965 (.02) | -.589  .559 (.16) | -.355  .723 (.10) |

Supplementary Table 2

*Confidence intervals (95%) for effect sizes over time for CMT and SPR tests in habitual contexts*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **L2+L1 group** | | |  |  | **L2-only group** | |  | |  | | **Control group** | | | | |
| Measure |  | CMT-Listen | CMT-Read | SPR | |  | CMT-Listen | CMT-Read | SPR | | |  | CMT-Listen | CMT-Read | SPR | | |
| Critical word | Whole sentence |  | Critical word | Whole sentence | |  | Critical word | | Whole sentence |
| *Matched contexts* |  |  | |  | |  |  | |  | | |  |  | |  | | |
| ***Pre-Post*** | LL, UL | -1.08, .28 | -1.02, .34 | .44, 1.90 | .32, 1.75 |  | -2.10,  -.61 | -.47, .88 | -.41, .95 | -.15, 1.22 | |  | -.57, .82 | -.37, 1.02 | -.45, .94 | -.09, 1.33 | |
| ***Pre-Delayed*** | LL, UL | -2.40, -.85 | -1.82, -.38 | .39, 1.84 | 17, 1.58 |  | -1.72,  -.29 | -1.10, .26 | -.25, 1.11 | .03, 1.36 | |  | -.57, .82 | -.08, 1.34 | -.42, .98 | -.06, 1.36 | |
| ***Post-Delayed*** | LL, UL | -1.42,  -.03 | -1.25, .12 | -.69, .65 | -.83, .51 |  | -.67, .67 | -.118, .19 | -.38, .97 | -.52, .83 | |  | -.69, .69 | -.44, .95 | -.64, .74 | -.72, .66 | |
| *Mismatched contexts* |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  | |
| ***Pre-Post*** | LL, UL | -1.04, .31 | -1.14, .22 | 15, 1.55 | -.14, 1.23 |  | -.50, .85 | -1.01,  .35 | -.57, .78 | .06, 1.46 | |  | -.42, .97 | -.81, .58 | -.52, .87 | -.22, 1.18 | |
| ***Pre-Delayed*** | LL, UL | -1.92, -.46 | -2.38, -.84 | .37, 1.81 | -.42, .93 |  | -.85, .50 | -.86,  .48 | -.52, .83 | -.11, 1.26 | |  | -.09, 1.54 | -.79, .60 | -.52, .87 | -.18, 1.22 | |
| ***Post-Delayed*** | LL, UL | -1.71, -.28 | -1.59, -.18 | -.34, 1.01 | -.93, .42 |  | -1.04, .32 | -.57,  .77 | -.59, .76 | -.81, .54 | |  | -.10, 1.32 | -.67, .71 | -.64, .75 | -.59, .80 | |

*Note.* CMT-Listen = Context-matching listening test, CMT-Read = Context-matching reading test, SPR = Self-paced reading test, LL = Lower limit, UL = Upper limit.

Negative effect sizes for the SPR indicate faster processing i.e. smaller reaction times.

Yellow highlighting indicates the effect sizes that do *not* pass through zero. Although not an infallible reflection of meaningful change (e.g. effect sizes that pass through zero can still be large, and/or statistically significant), the observed pattern of such confidence intervals gives further indication of the general pattern of findings, showing more benefits across the different measures and trial types for the L2+L2 group than the L2-only group, and more in both treatment groups compared to the Control group.

Supplementary Table 3

*Confidence intervals (95%) for effect sizes over time for CMT and SPR tests in ongoing contexts*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **L2+L1 group** | | |  |  | **L2-only group** | |  | |  | | **Control group** | | | | |
| Measure |  | CMT-Listen | CMT-Read | SPR | |  | CMT-Listen | CMT-Read | SPR | | |  | CMT-Listen | CMT-Read | SPR | | |
| Critical word | Whole sentence |  | Critical word | Whole sentence | |  | Critical word | | Whole sentence |
| Matched contexts |  |  | |  | |  |  | |  | | |  |  | |  | | |
| **Pre-Post** | LL, UL | -1.21, .16 | -2.40,  -.85 | .59, 2.08 | .46, 1.92 |  | -1.42,  -.04 | -1.71,  -.29 | .25,  1.67 | -.48,  .86 | |  | -.89,  .50 | -.89,  .50 | -.09,  .133 | -.16,  1.25 | |
| **Pre-Delayed** | LL, UL | -2.46, .90 | -2.65,  -1.04 | .45, .190 | .30, 1.73 |  | -1.04,  .32 | -1.77,  -.34 | -.34,  1.01 | -.17,  1.20 | |  | -.48,  .91 | -.85,  .54 | -.24,  1.16 | -.18,  1.23 | |
| **Post-Delayed** | LL, UL | -1.40, .01 | -1.01, .34 | -.76, .59 | -.82, .53 |  | -.44,  .91 | -.67,  .67 | -1.22,  .15 | -.30,  1.06 | |  | -.27,  1.13 | -.64,  .74 | -.71,  .68 | -.78,  .61 | |
| Mismatched contexts |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  | |
| **Pre-Post** | LL, UL | -1.09, .27 | -1.61,  -.20 | .27, 1.70 | -.20, 1.16 |  | -1.14,  .22 | -.86,  .49 | -.48,  .87 | -.18,  1.19 | |  | -.41,  .98 | -1.08, .32 | -.58,  .81 | -.43,  .97 | |
| **Pre-Delayed** | LL, UL | -1.98,  -.51 | -2.18,  -.67 | .25, 1.67 | -.40, .95 |  | -.81,  .53 | -.98,  .38 | -.49,  .86 | -.29,  1.07 | |  | -.02,  1.41 | -.95,  .44 | -.79,  .60 | -.20,  1.21 | |
| **Post-Delayed** | LL, UL | -1.38,  -.00 | -1.10, .26 | -.72, .63 | -.84, .51 |  | -.33,  1.03 | -.76,  .59 | -.66,  .69 | -.77,  .57 | |  | -.31,  1.09 | -.61,  .78 | -.87,  .52 | -.53,  .86 | |

*Note.* CMT-Listen = Context-matching listening test, CMT-Read = Context-matching reading test, SPR = Self-paced reading test, LL = Lower limit, UL = Upper limit

Negative effect sizes for the SPR indicate faster processing i.e. smaller reaction times.

Yellow highlighting indicates the effect sizes that do *not* pass through zero. Although not an infallible reflection of meaningful change (e.g. effect sizes that pass through zero can still be large, and/or statistically significant), the observed pattern of such confidence intervals gives further indication of the general pattern of findings, showing more benefits across the different measures and trial types for the L2+L2 group than the L2-only group, and more in both treatment groups compared to the Control group.

Supplementary Table 4

*Normality tests for CMT tests*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| PRE\_MISMATCHHAB\_LISTEN | .138 | 50 | .018 | .933 | 50 | 0.04 |
| POST\_MISMATCHHAB\_LISTEN | .121 | 50 | .066 | .949 | 50 | 0.03 |
| DELAYED\_MISMATCHHAB\_LISTEN | .136 | 50 | .021 | .937 | 50 | 0.01 |
| PRE\_MATCHHAB\_LISTEN | .097 | 50 | .200\* | .941 | 50 | 0.01 |
| POST\_MATCHHAB\_LISTEN | .198 | 50 | .000 | .858 | 50 | 0.00 |
| DELAYED\_MATCHHAB\_LISTEN | .225 | 50 | .000 | .819 | 50 | 0.00 |
| PRE\_MATCHONG\_LISTEN | .133 | 50 | .028 | .917 | 50 | 0.00 |
| POST\_MATCHONG\_LISTEN | .192 | 50 | .000 | .884 | 50 | 0.00 |
| DELAYED\_MATCHONG\_LISTEN | .210 | 50 | .000 | .858 | 50 | 0.00 |
| PRE\_MISMATCHONG\_LISTEN | .136 | 50 | .021 | .951 | 50 | 0.04 |
| POST\_MISMATCHONG\_LISTEN | .226 | 50 | .000 | .730 | 50 | 0.00 |
| DELAYED\_MISMATCHONG\_LISTEN | .104 | 50 | .200\* | .939 | 50 | 0.01 |

Supplementary Table 5

*Normality tests for SPR tests*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| PRETEST\_MATCH\_HAB\_SPR | .171 | 50 | .001 | .868 | 50 | 0.00 |
| POST\_MATCH\_HAB\_SPR | .129 | 50 | .037 | .943 | 50 | 0.02 |
| DELAYED\_MATCH\_HAB\_SPR | .148 | 50 | .008 | .816 | 50 | 0.00 |
| PRETEST\_MATCH\_ONG\_SPR | .091 | 50 | .200\* | .942 | 50 | 0.02 |
| POST\_MATCH\_ONG\_SPR | .176 | 50 | .001 | .790 | 50 | 0.00 |
| DELAYED\_MATCH\_ONG\_SPR | .110 | 50 | .176 | .937 | 50 | 0.01 |
| PRETEST\_MISMATCH\_HAB\_SPR | .195 | 50 | .000 | .893 | 50 | 0.00 |
| POST\_MISMATCH\_HAB\_SPR | .167 | 50 | .001 | .900 | 50 | 0.00 |
| DELAYED\_MISMATCH\_HAB\_SPR | .206 | 50 | .000 | .724 | 50 | 0.00 |
| PRETEST\_MISMATCH\_ONG\_SPR | .124 | 50 | .055 | .904 | 50 | 0.00 |
| POST\_MISMATCH\_\_ONG\_SPR | .089 | 50 | .200\* | .945 | 50 | 0.02 |
| DELAYED\_MISMATCH\_\_ONG\_SPR | .135 | 50 | .023 | .897 | 50 | 0.00 |
| MATCH\_HAB\_WHOLE\_SPR\_PRETEST | .151 | 50 | .006 | .925 | 50 | 0.00 |
| MATCH\_HAB\_WHOLE\_SPR\_POST | .143 | 50 | .012 | .910 | 50 | 0.00 |
| MATCH\_HAB\_WHOLE\_SPR\_DELAYED | .173 | 50 | .001 | .842 | 50 | 0.00 |
| MATCH\_ONG\_WHOLE\_SPR\_PRETEST | .126 | 50 | .007 | .972 | 50 | 0.00 |
| MATCH\_ONG\_WHOLE\_SPR\_POST | .149 | 50 | .010 | .972 | 50 | 0.00 |
| MATCH\_ONG\_WHOLE\_SPR\_DELAYED | .089 | 50 | .200\* | .984 | 50 | 0.04 |
| MISMATCH\_HAB\_WHOLE\_SPR\_PRETEST | .254 | 50 | .000 | .859 | 50 | 0.00 |
| MISMATCH\_HAB\_WHOLE\_SPR\_POST | .061 | 50 | .200\* | .887 | 50 | 0.03 |
| MISMATCH\_HAB\_WHOLE\_SPR\_DELAYED | .170 | 50 | .001 | .876 | 50 | 0.00 |
| MISMATCH\_ONG\_WHOLE\_SPR\_PRETEST | .263 | 50 | .000 | .594 | 50 | 0.00 |
| MISMATCH\_ONG\_WHOLE\_SPR\_POST | .131 | 50 | .033 | .917 | 50 | 0.00 |
| MISMATCH\_ONG\_WHOLE\_SPR\_DELAYED | .141 | 50 | .015 | .811 | 50 | 0.00 |
| \*. This is a lower bound of the true significance. | | | | | | |
| a. Lilliefors Significance Correction | | | | | | |