**Appendix A. Classifiers and Nouns**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Japanese Classifier | Noun in Japanese | English translation | Chinese Classifier | Noun in Chinese |
| 冊 | 絵本 | picture-book | 本/册 | 绘本 |
| 冊 | 雑誌 | magazine | 本 | 杂志 |
| 冊 | 辞書 | dictionary | 本 | 字典 |
| 冊 | ノート | notebook | 本 | 笔记本 |
| 匹 | 魚 | fish | 条 | 鱼 |
| 匹 | 猿 | monkey | 只 | 猴子 |
| 匹 | 犬 | dog | 只 | 狗 |
| 匹 | ネズミ | mouse | 只 | 老鼠 |
| 台 | ピアノ | piano | 架 | 钢琴 |
| 台 | ベッド | bed | 张 | 床 |
| 台 | カメラ | camera | 架/个 | 相机 |
| 台 | 携帯電話 | mobile phone | 个 | 手机 |
| 枚 | ハンカチ | handkerchief | 块 | 手帕 |
| 枚 | シャツ | shirt | 件 | 恤衫 |
| 枚 | 葉っぱ | leaves | 片 | 叶 |
| 枚 | 鏡 | mirror | 面 | 镜子 |
| 着 | 服 | cloth | 套 | 穿着 |
| 着 | ドレス | dress | 条/件 | 裙子 |
| 着 | コート | coat | 件 | 外套 |
| 着 | スーツ | suit | 套 | 诉讼 |
| 羽 | カラス | crow | 只 | 乌鸦 |
| 羽 | ニワトリ | chicken | 只 | 鸡 |
| 羽 | 鳥 | bird | 只 | 鸟 |
| 羽 | はと | pigeon | 只 | 鸽子 |
| 足 | スリッパ | slippers | 双 | 拖鞋 |
| 足 | ブーツ | boots | 双 | 靴子 |
| 足 | 靴下 | socks | 双 | 袜子 |
| 足 | サンダル | sandals | 双 | 凉鞋 |
| 軒 | アパート | apartment | 栋 | 公寓 |
| 軒 | コンビニ | convenience store | 个/间 | 便利店 |
| 軒 | スーパー | supermarket | 摊/店 | 超市 |
| 軒 | 居酒屋 | Japanese tavern | 个/间 | 酒馆 |

**Appendix B. Visual-World Task: Classifiers and *Ni*/*De***

**Classifiers**

As shown in Figure 6, each display for classifiers consisted of a Target noun (e.g., two dresses), a Competitor noun (e.g., two picture books) and two other distractors. The number of the target and competitor nouns was always equal (varying from 1 to 4), and the number of distractor objects was always different from both of them. Two types of critical trials were created: (3) classifier-matched trials and (4) classifier-mismatched trials.

[INSERT FIGURE 6 AROUND HERE]

The critical sentence in both trials always formed the following order: Number-Classifier-Genitive-[relative clause]-Noun. Across trials, one of the classifier-noun pairs was defined as a Target (e.g., two dresses), whereas the other pair was defined as a Competitor. The classifier (e.g., *chaku*;a counter for clothes) was matched to the target word (e.g., dresses) in the classifier-matched trials (defined as Target trials), whereas the other classifier (e.g., *satsu*; a counter for books) was matched to the competitor noun (e.g., picture books) in the classifier-mismatched trials (defined as Competitor trials). The eye-movements were analyzed from the onset of a classifier. If participants can use the information of the classifier in real time, then the looks to the Target (e.g., dresses) will be greater in the Target trials (*chaku*) than in the Competitor trials (*satsu*).

**Ni/De**

As shown in Figure 7, each display for classifiers consisted of person A (performing an action in a location), person B (getting to the same location as person A), and two distractor objects. The location depicted in the pictures was the same for person A and person B (e.g., train). Person A is defined as a Target, whereas person B is defined as a Competitor. Two types of critical trials were created: (5) *de* trials and (6) *ni* trials.

[INSERT FIGURE 7 AROUND HERE]

The first part of the sentence always formed the following pattern: NP-*de*/*ni*-Adverb-VP. The action of person A indicated motion toward the location (e.g., get on) in the *ni* trials (defined as Target trials), whereas the action of person B took place at the location (e.g., talk over the phone) in the *de* trials (defined as Competitor trials). The eye-movements were analyzed from the onset of the particle, *ni* or *de*. If participants were sensitive to the locative particle, then the looks to the Target would be greater in the Target trials (*de*) than in the Competitor trials (*ni*).

**Appendix C. Analyses of the Visual-World Task**

The sensitivity index was computed as TA difference scores (TA in the Target trials – TA in the Competitor Trials) during the time-locked window (i.e., 200 ms[[1]](#footnote-1) from the data-driven onset). We identified the “data-driven” onset where Japanese NSs actually used the linguistic trigger to differentiate the Target from the Competitor in the display. Since it was unknown exactly when listeners started to use the cues from each of the three target grammatical structures, the data-driven onset is a valid point for determining the onset of eye-movement analysis for computing the sensitivity index. TAs were plotted separately for the three target structures from the onset of the target linguistic trigger (i.e., *ga* or *o*, classifier, and *ni* or *de*). The data-driven onset was identified by comparing the TA in the Target trials and the Competitor trials; it was the time point where TAs in the Target trials became a standard error larger than those in the Competitor trials. The post-hoc region was set based on the native speakers’ data, not on L2 learners because the current study attempted to assess L2 implicit knowledge that was qualitatively similar to NSs.

The computation procedure of sensitivity index for the visual-world task is elaborated by an illustrative example of the transitive/intransitive structure. First, TA scores were computed separately for the transitive trials and the intransitive trials (see Figure 8). Here, the TA scores referred to the proportion of looks to the Target (e.g., the mother) over the total looks of the Target and the Competitor (e.g., the table). The TA scores were higher in the transitive trials than in the intransitive trials at 500 ms after the linguistic trigger (i.e., ga or o). This means that the participants started to use the linguistic cues to distinguish the Target and the Competitor at 500 ms. The data-driven onset of fixations was thus set at 500 ms. The sensitivity index was then computed by subtracting the TA scores in the intransitive trials from the TA scores in the transitive trials over the critical region (i.e., from 500 ms to 700 ms). The higher the score, the more likely that participants looked at the Target rather than the Competitor in the transitive trials compared to in the intransitive trials.

 [INSERT FIGURE 8 AROUND HERE]

The results for the classifier and the *ni*/*de* are also presented below. TA scores in the classifier construction are illustrated in Figure 9. An equal proportion of looks was observed at 0 ms between the classifier-match trials (target) and the classifier-mismatch trails (competitor). Native speakers started to look more at the Target in the classifier-match trials than in the classifier-mismatch trials 100 ms after the linguistic trigger (i.e., a classifier). The data-driven onset of fixations was thus set at 100 ms for the classifier construction.

[INSERT FIGURE 9 AROUND HERE]

TA scores in the *ni/de* construction are illustrated in Figure 10. An equal proportion of looks was observed at 0 ms between the *de* trials (target) and the *ni* trials (competitor). Native speakers started to look more at the Target in the *de* trials than in the *ni* trials 100 ms after the linguistic trigger (i.e., *ni* or *de*). The data-driven onset of fixations was thus set at 100 ms for the *ni*/*de* construction.

[INSERT FIGURE 10 AROUND HERE]

**Appendix D. Stimulus Sentences for the Word-Monitoring Task**

|  |  |
| --- | --- |
| Condition | Stimulus Sentences |
| Transitive | *Ao to kiiro no enogu o/\*ga* ***mazeru*** *to, kirei na mimdori ni naru.* |
| Blue and yellow paint-ACC/SUB mix if, beautiful green become |
| When you mix blue and yellow paint, it becomes beautiful green. |
| Intransitive | *Ao to kiiro no enogu ga/\*o* ***mazaru*** *to, kirei na mimdori ni naru.* |
| Blue and yellow paint-ACC/SUB mix if, beautiful green become |
| When you mix blue and yellow paint, it becomes beautiful green. |
| Classifier | *Kumiko-san-wa san-dai/\*satsu no* ***keitaidenwa*** *o motte imasu.* |
| *Kumiko-TOPIC three-CLASSIFIER GEN cell phone-OBJ have is* |
| Kumiko has three cellphones. |
| Ni/De | *Atatakai toki ni soto de/\*ni* ***neru*** *to kimochi ii.* |
| Warm time outside-LOC sleep if comfortable |
| It is comfortable to sleep outside in the warm weather. |

Note. The monitoring words are bold and underlined. For the transitive structures, the target sentence always included a segment of the case-marking particle (*ga* or *o*) and a verb (transitive or intransitive). The target word was always the verb following the particles (*ga* or *o*)*.* For classifiers, the classifier-noun pairs were embedded in a carrier sentence. The target word was the noun that followed the classifier and the genitive particle *no*. For the *ni/de* construction, a segment of a location noun, the particle (*de/\*ni*), and the action verb was always included in the sentence.[[2]](#footnote-2) The target monitoring word was the verb following *ni* or *de* (i.e., *neru*).

**Appendix E.** **Stimulus Sentences for the Self-Paced Reading Task**

|  |  |
| --- | --- |
| Condition | Stimulus Sentences |
| Transitive | *Uta no gurupu o/ tsukuru tokini/ danshi to/ joshi o(\*ga)/* ***mazeru to/ ii****/ baransu ni/ naru to omou/*Singing group-OBJ/ make when boy and girls-OBJ mix if good balance becomes thinkWhen you form a singing group, I think it makes a good balance if you mix boys and girls. |
| Intransitive | *Uta no gurupu o/ tsukuru tokini/ danshi to/ joshi ga(\*o)/* ***mazaru to/ ii****/ baransu ni/ naru to omou/*Singing group-OBJ/ make when boy and girls-SUB mix if good balance becomes thinkWhen you form a singing group, I think it makes a good balance if boys and girls are mixed. |
| Classifier | *Kono/ omise de/ juu-dai (\*mai)-no/* ***keitaidenwa ga/ nusumareta to/*** *keisatsu ga/ houkoku o/ uketa soudesu.*This shop-LOC ten-CLASSIFER GEN cell phone-SUB stolen that policie-SUB report-OBJ received heardI heard that the police received the report that ten cellphones were stolen at this shop. |
| Ni/De | *Kinou ane to/ ohiru no / san-ji goro/ resutoran de (\*ni)/* ***shokujishinagara,/ ryokou no/*** *hanashi o/ shimashita.*Yesterday sister-with noon-GEN three about restaurant-LOC eat while trip-GEN talk didI talked about a trip with my elder sister at the restaurant [at] about 3 in the afternoon. |

Note. The critical regions are bold and underlined.

**Appendix F.Stimulus Sentences for the timed SPOT**

 (1) Transitive/Intransitive

パスワードを知らないまま金庫○閉めたので、金庫から宝石が取り出せません。

*Password wo shiranai mama kinko ○ shimeta node, kinko kara houseki ga toridasemasen*.

Password-OBJ without knowing safe-OBJ close because safe from jewel-SBU can’t take out

(2) Classifiers

この町の消防署の隣には、２○○のコンビニが並んで建っている。

*Kono machi no shoubousho no tonari niwa ni-○○ no konbini ga narande tatte iru.*

This town-GEN fire station-GEN beside 2-Classisifier GEN convenience store-SUB built

(3) *Ni/De*

多くの高校生は友達とカラオケボックス○歌って踊ります。

*Ooku no koukousei wa tomodachi to karaokebokkusu ○* *utatte odorimasu.*

Many high schoolers-SUB friends with karaoke box-LOC sing and dance

**Appendix G. Data Cleaning prodcedures and results of eye-movement and RT data**

**Visual-World task**

To ensure that the participants were focusing on meaning, accuracy of the comprehension questions was computed. The mean accuracy score was 90.83 (SD = 4.12) for the L2 learners. Missing frames due to blinking or looks outside of the screen accounted for 9.4% in the L2 learners. For all remaining frames, location of fixation was coded as a look towards one of the quadrants.

**Word-Monitoring Task**

The mean accuracy score for the comprehension questions was 91.55 (*SD* = 5.14) for the L2 learners. In order to screen the RT data, outliers were discarded that fell outside the low and high cutoffs set at 100 ms and 2500 ms, or that were 3 SDs above or below each participant’s mean, respectively. The higher cutoff was set in order to exclude responses in which participants inadvertently forgot to respond to the target word, and the lower cutoff was set to exclude the responses given without hearing a target word. These procedures, along with display errors, eliminated 3.3% of the data.

**Self-Paced Reading Task**

The mean accuracy scores of the comprehension questions were 91.55 (*SD* = 4.82) for the L2 learners. In order to screen the RT data, outliers were discarded that fell outside the low and high cutoffs set at 120 ms and 1500 ms or that were 3 SDs above or below each participant’s mean, respectively. The higher cutoff was set in order to exclude responses in which participants were reading too slowly, and the lower cutoff was set to exclude the responses given without reading the words. These procedures, along with display errors, eliminated 6.6% of the data for L2 learners.

**Appendix H. Results of *T*-Tests and Descriptive Statistics for the Language Tests by Native Speakers**

 Descriptive statistics for the six language tests performed by native speakers are summarized in the table below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|   | *N* | Possible Max | *M* | *SD* | Min | Max | 95% CI |
| Eyea | 20 | - | 0.08 | 0.14 | -0.18 | 0.37 | [0.01, 0.14] |
| WMa | 24 | - | 105 | 39 | 6 | 193 | [89, 122] |
| SPRa | 16 | - | 35 | 61 | -81 | 169 | [2, 67] |
| T-AGJTb | 31 | 100 | 91.52 | 6.21 | 68.75 | 100 | [0.89, 0.94] |
| T-VJGTb | 31 | 100 | 90.66 | 6.9 | 74.11 | 100 | [0.88, 0.93] |
| T-SPOTb | 31 | 100 | 90.59 | 17.55 | 18.75 | 100 | [0.84, 0.97] |

a. The values for the online comprehension tasks indicate sensitivity index.

b. The values for the form-focused tasks indicate percentage accuracy score.

Note 1. Eye = Visual-World task, SPR = Self-Paced Reading task, WM = Word-Monitoring task, T-AGJT = Timed Auditory GJT, T-VGJT = Timed Visual GJT, T-SPOT = Timed SPOT, CI = Confidence Interval.

Note 2. A group of thirty-one participants took the entire test battery. After the initial testing, we found that the visual-world task did not work as expected. Another group of 20 participants was tested with only the revised visual-world task, which was eventually administered to all the L2 learners. The numbers of participants in the word-monitoring and self-paced reading tasks were different because some test items were revised and only the data from the finalized tasks were included.

 For the native speakers, the target advantage (TA) scores in the Target trials were significantly greater than those in the Competitor trials, supporting that they were sensitive to the linguistic manipulation of the task.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Target Advantage Score in Target Trials |  | Target Advantage Score in Competitor Trials |  |  |  |  |  |
|   | *M* | *SD* |   | *M* | *SD* | *t* | *df* | *p* | Cohen's *d* | 95% CI |
| Visual-world | .55 | .11 |   | .47 | .06 | 2.438 | 19 | .02 | 0.56 | [0.01, 0.14] |

 For the word-monitoring and self-paced reading tasks, native speakers’ RTs were significnalty longer in the ungrammatical items than in the grammatical items. In other words, as they registered the grammatical errors online, they slowed down significantly to respond to the ungrammatical items compare to the grammatical items.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Reaction Time in Grammatical Items |  | Reaction Time in Ungrammatical Items |  |  |  |  |  |
|   | *M* | *SD* |   | *M* | *SD* | *t* | *df* | *p* | Cohen's *d* | 95% CI |
| Word-monitoring | 399 | 66 |  | 504 | 83 | 13.099 | 23 | .00 | 2.94 | [89, 122] |
| Self-paced reading | 575 | 161 |   | 610 | 186 | 2.250 | 15 | .04 | 0.79 | [2, 67] |

**Appendix I. Results of *T*-Tests and Descriptive Statistics for the Language Tests by L2 Learners**

Unlike the native speakers, in the L2 learners’ group, the target advantage (TA) scores in the Target trials were not significantly different from those in the Competitor trials. The effect size was smaller for the L2 learners than that for the native speakers (see Appendix H above).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | Target Advantage Score in Target Trials |   | Target Advantage Score in Competitor Trials |   |   |   |   |   |
|   | *M* | *SD* |   | *M* | *SD* | *t* | *df* | *p* | Cohen's *d* | 95% CI |
| Visual-world | 0.5 | 0.07 |   | 0.49 | 0.07 | 1.433 | 99 | 0.16 | 0.14 | [-0.01, 0.03] |

For the word-monitoring and self-paced reading tasks, L2 learners’ RTs were significnalty longer in the ungrammatical items than in the grammatical items. The same magnititude of effect size was observed for the two tasks, but it was smaller than the native speakers’ group (see Appendix H above).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Reaction Time in Grammatical Items |  | Reaction Time in Ungrammatical Items |  |  |  |  |  |
|   | *M* | *SD* |   | *M* | *SD* | *t* | *df* | *p* | Cohen's *d* | 95% CI |
| Word-monitoring | 493 | 124 |  | 515 | 129 | 4.056 | 99 | .00 | 0.41 | [11, 33] |
| Self-paced reading | 1086 | 256 |   | 1122 | 279 | 3.978 | 99 | .00 | 0.41 | [18, 54] |

 **(3) Classifier-matched trials (Target trials)**

*Ni-chaku no naraberareta doresu ga kono heya ni arimasu.*

Two-CHAKU GEN laid out dress-SUB this room-LOC exist.

(There are two dresses in this room)

**(4) Classifier-mismatched trials (Competitor trials)**

*Ni-satsu no naraberareta ehon ga kono heya ni arimasu.*

Two-SATSU GEN laid out picture book-SUB this room-LOC exist.

(There are two picture books in this room)

*Figure 6*. Visual Scene and Critical Sentences for Classifier

**(5) *De* trials (Target trials)**

*Otoko no hito ga densha de katte ni denwa wo shite imasu.*

Man SUB train-LOC selfishly calling is

(The man is talking selfishly on the phone on the train.)

**(6) *Ni* trials (Competitor trials)**

*Otoko no hito ga densha ni katte ni norimashita.*

Man SUB train-LOC selfishly got on

 (The man got on the train selfishly.)

*Figure 7*. Visual Scene and Critical Sentence for *Ni*/*De*

*Figure 8.* Target Advantage in Transitive Trials and Intransitive Trials (Native Speakers, *n* = 20)

*Figure 9.* Target Advantage in Classifier-matched Trials and Classifier-mismatched trials (Native Speakers, *n* = 20)

*Figure 10.* Target Advantage in De Trials and Ni Trials (Native Speakers, *n* = 20)

1. We focused on the first 200 ms time windows for computing the sensitivity index for the following two reasons. First, implicit knowledge should be deployed very rapidly. It is assumed that setting a longer time window might increase the chance of including the use of explicit knowledge. Explicit knowledge was more likely to be developed when the inter-responses took longer than 200 ms in the domain-general implicit learning task (Destrebecqz, A. and A. Cleeremans (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, 8(2): 343-350.). Second, reaction-time tasks, particularly the word-monitoring task, are pre-time-locked, in order for the online sensitivity to be revealed. The errors were embedded right before the target word, and error detection should happen almost at the exact point in time where ungrammatical segments occur in speech. As the first justification explains, it is important to capture the earliest sensitivity to the target linguistic structures. [↑](#footnote-ref-1)
2. Due to the constraints in the design of the visual-world task, the comparison of *ni* and *de* in the word-monitoring task and the self-paced reading task was slightly different from that in the visual-world task. In the visual-world task, the looks to the picture involving action (compatible with *de*) and those to the picture involving motion (compatible with *ni*) were compared. In the word-monitoring and self-paced reading tasks, the reaction time to the action verb (compatible with *de*) and that to the static verb (compatible with *ni*) were compared. The baseline is different (*ni* motion and static verbs), but the critical target structure (*de* for action verbs) is the same across the tasks. [↑](#footnote-ref-2)