Supplementary Material:
The Interaction of Visual and Linguistic Saliency during Syntactic Ambiguity Resolution

Moreno I. Coco and Frank Keller
Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, UK
Phone: +44 131 650 4407, Fax: +44 131 650 4587
mcoco@staffmail.ed.ac.uk, keller@inf.ed.ac.uk

Decay of Visual Saliency Effect During Preview

It is important to make sure that the anticipatory effects of visual saliency observed at the verb phrase in Experiments 1 and 3 are due to an interaction of saliency with sentence processing, rather than being purely visual. We therefore examine whether early effects of visual saliency decay during the preview time. If a pure effect of visual saliency persists during the whole trial, then we should observe fixation to the salient object to be steady across the preview time. If instead visual saliency has an early effect and then decays prior to the sentence onset, then effects observed during linguistic processing can be attributed to an interaction of visual saliency with linguistic information concurrently processed.

For this comparison, we utilize the eye-movement data from Experiment 3, where the preview time was 1000 ms across all participants. In particular, we consider from scene onset until the sentence onset (from 0 to 1000 ms, 100 ms windows) and compute fixation probability to the visually salient object (e.g., Single Location), and contrast this with the other object whose saliency was not manipulated (e.g., Compound Location), and with the rest of the objects in the display. We perform this analysis separately for the two conditions of visual saliency (Single Location and Compound Location), as they differ in their complexity and plausibility. The data from the two Intonational Break conditions can be collapsed, as it does not intervene during the preview. Fixation probability is modeled as a function of Saliency (see Table 1 for details on the coding of the factors) and Time (as a polynomial of order two) using linear mixed effect models (see section Analysis in the main text for details).

In Figure 1, it is immediately clear that there is an effect of visual saliency at the onset of the trial, with more looks to the salient object than to the non-salient one. This is true for both Single Location (left panel) and Compound Location (right panel), where the effect sets in later. These results are inferentially confirmed in the LME models, reported in Table 1 for the Single Location and Table 2 for the Compound Location.

Quaterly Journal of Experimental Psychology 68: 1, 46–74, 2015.

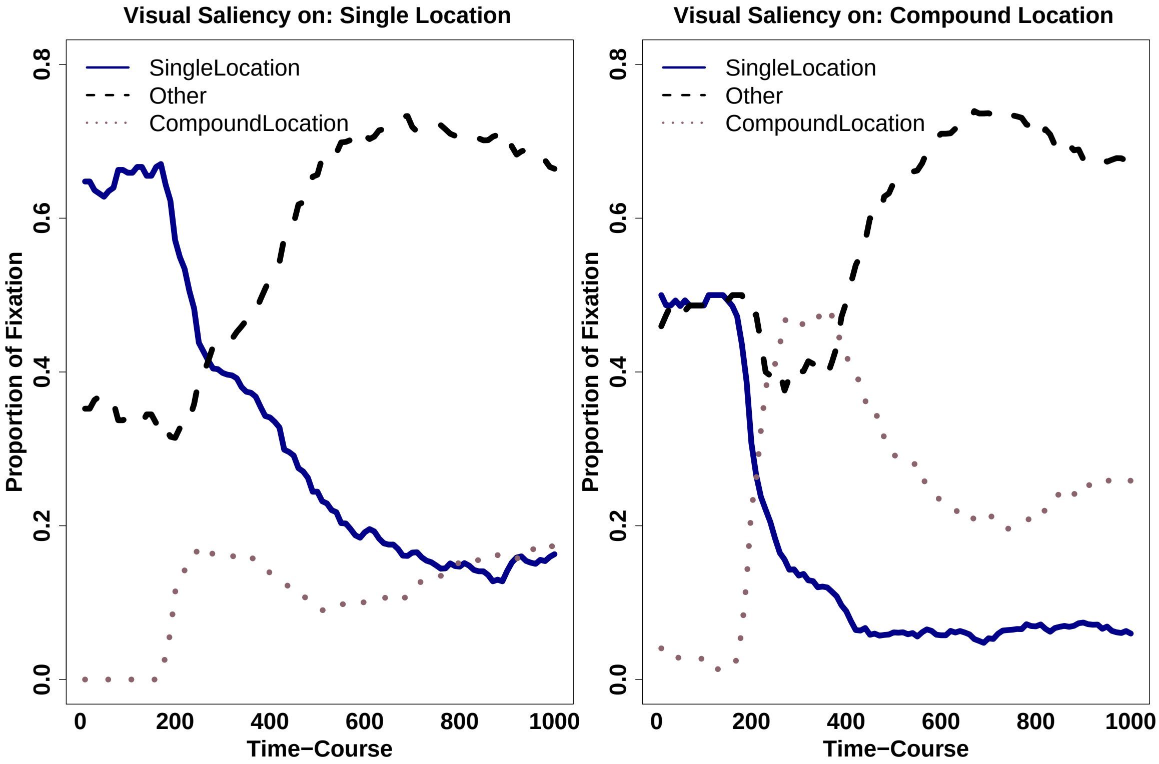


Figure 1. Effects of visual saliency at preview. Time-course plot of the probability of fixations on the *Single Location*, *Compound Location* and *Other* objects of the visual context (excluding the white background) from the onset of the scene until the onset of the sentence (0 to 1000 ms). Left panel: visual saliency on Single Location. Right panel: visual saliency on Compound Location. The data is taken from Experiment 3, where preview time was fixed at 1000 ms for all participants. The data from the two intonational break conditions has been aggregated, as the stimuli are identical during preview.

Table 1

Effects of visual saliency at preview (data from Experiment 3). Mixed model analysis of the probability of fixations on the Object*. Single Location was coded as* 0.5*, Other as* −0.5*. Time (0 to 1000 ms, in 100 ms intervals) is represented as an orthogonal polynomial of order two (Time*^{1}*, Time*^{2}*).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.104	0.016	6.547	<1e-04
Object	0.087	0.015	5.674	<1e-04
Time^1	-0.032	0.042	-0.763	0.4
Time^2	-0.048	0.062	-0.787	0.4
Object:Time^1	-0.01	0.015	-0.676	0.5
Object:Time^2	-0.102	0.015	-6.828	<1e-04

Table 2

Effects of visual saliency at preview (data from Experiment 3). Mixed model analysis of the probability of fixations on the Object. *Other was coded as* 0.5, *Compound Location as* -0.5. *Time (0 to 1000 ms, in 100 ms intervals) is represented as an orthogonal polynomial of order two (Time*^{1}*, Time*^{2}*).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.118	0.014	8.438	<1e-04
Object	-0.093	0.014	-6.516	<1e-04
Time^1	0.111	0.045	2.452	0.01
Time^2	-0.103	0.020	-4.962	<1e-04
Object:Time^1	-0.011	0.016	-0.672	0.5
Object:Time^2	0.152	0.016	9.219	<1e-04

We observe a significantly higher probability of fixating on the salient object than on any other object in the display. However, this fixation probability decreases over time, as indicated by the significant two-way interaction Object:Time^2. Looks to the salient object sharply increase after scene onset, and then decrease as the preview ends. This effect indicates a shift of attention that is guided by a bottom-up, low-level scene information, and is not related to the linguistic input, as predicted by models of visual saliency (see Itti & Koch, 2000). However, visual attention is progressively allocated to the non-salient Other objects of the visual display, and the likelihood of fixating a salient object drops at the onset of the sentence. Top-down control starts to act on visual attention, shifting attention away from the salient object.

Experiment 1: Referentiality Effect at ROI:NP *the orange* on Object ORANGE

Figure 2 compares the probability of fixations on the target object ORANGE (2 Referents) or distractor (1 Referent) across the six different experimental conditions of Experiment 1. Note that the orange and the distractor are the same target object for the purposes of our analysis, as the single orange is replaced with by the distractor in the 1 Referent condition.

We observe a trend of more looks to the target object in the 2 Referents condition, which marginally increases with time. However, Number of Referent is not statistically significant as a main effect, nor in the two-way interaction with time. The effect of two referents becomes significant only in the three-way interaction with time and visual saliency. In particular, when the Single Location is salient, looks to the target increase steeply, especially for the 2 Referent condition. This result corroborates with the analysis of visual saliency presented in Table ... of the main text.

Before the direct object *the orange* is completely spelled out, salient objects attract visual attention. The saliency activation triggers a visual competition with the mentioned object. However, once the linguistic referent has been fully perceived and saliency has exhausted its anticipatory potential, attention is allocated to the direct object. This effect, however, emerges only for the two referent condition.

This result partially aligns with previous work on syntactic resolution in the visual world paradigm, (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), as we also found evidence for visual competition in the two-referent context, in which two visual objects share the same

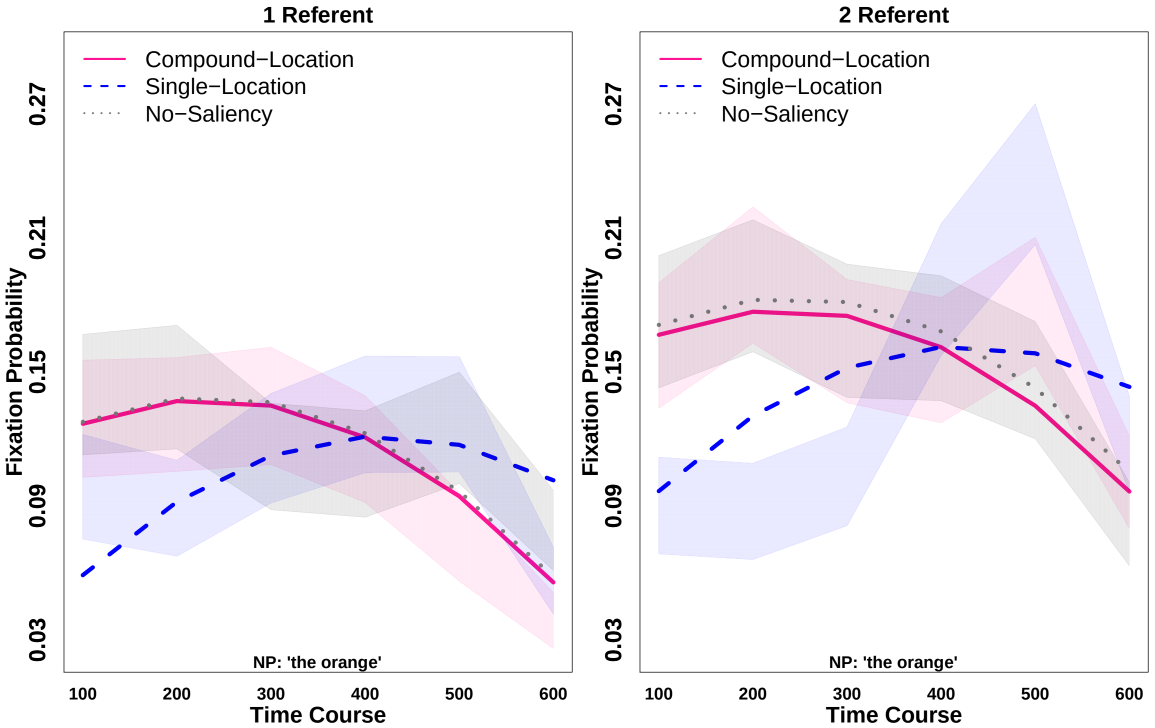


Figure 2. Experiment 1: Time course plot of fixation probability for the object ORANGE (right panel, for the 2 Referent condition) or DISTRACTOR (left panel, for the 1 Referent condition, where the only orange in the display was depicted on the tray) at ROI:NP *the orange* for Long and Short preview collapsed. The saliency conditions (No Saliency, Single Location, and Compound Location) are marked through line types and colors. The shaded bands indicate the standard error around the observed mean. The lines represent the predicted values of the LME model reported in Table 3. Note that the offset of the region of analysis varied by items, but fixations crossing the offset were excluded, see Analysis section for details.

linguistic referent: at *the orange*, a single ORANGE and an ORANGE ON TRAY compete for attention. However, we extend this previous work by showing that the effect is modulated by the visual saliency of alternative objects available in the visual context. In fact, when saliency is manipulated, we find that visual competition in the two-referent condition is delayed, presumably because visual attention is allocated to evaluate whether the salient object is going to be the direct object of verb *put*.

Experiment 2: Linguistic Prominence Effect on ROI:1PP *on the tray* on Object ORANGE ON TRAY

Intonational breaks have a similar effect also on the other compound object ORANGE ON TRAY, but this time for the other intonation, i.e., NP-modifier, see Figure 3. This is again reflected in a main effect of Intonation: looks to the target object are higher for the NP modifier condition than for the PP modifier condition, see Table 4. Moreover, fixation probabilities pattern as a parabola over time, as highlighted by the interaction of Intonation with both the linear and quadratic terms

Table 3

Experiment 1. Mixed model analysis of the fixations on the object ORANGE *(2 Referent condition) or* DISTRACTOR *(1 Referent condition) at ROI:NP* the orange. Saliency *is contrast coded with* No Saliency *as reference level,* Number of Referents *is coded as* -0.5 *for 1 Referent, 0.5 for 2 Referents. Time is represented as an orthogonal polynomial of order two (Time^1, Time^2).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.125	0.350	0.359	0.7
SaliencyCompoundLoc	0.008	0.024	0.358	0.7
Time^2	-0.036	0.010	-3.438	0.0001
SaliencySingleLoc	-0.018	0.021	-0.849	0.3
Time^1	-0.023	0.018	-1.259	0.2
Referent	0.044	0.082	0.537	0.5
SaliencyCompoundLoc:Referent	0.026	0.019	1.339	0.1
Time^1:Referent	0.029	0.019	1.554	0.1
Time^2:Referent	-0.011	0.019	-0.627	0.5
SaliencySingleLoc:Time1:Referent	0.118	0.046	2.536	0.01

Formula: (1 | item) + (1 | participant) + SaliencyCompoundLoc + (0 + SaliencyCompoundLoc | participant) + (0 + SaliencyCompoundLoc | item) + Time^2 + (0 + Time^2 | item) + SaliencySingleLoc + (0 + SaliencySingleLoc | participant) + (0 + SaliencySingleLoc | item) + Time^1 + (0 + Time^1 | item) + (0 + Time1 | participant) + Referent + (0 + Referent | item) + (0 + Referent | participant) + SaliencyCompoundLoc:Referent + Referent:Time^1 + Referent:Time^2 + SaliencySingleLoc:Referent:Time^1

Table 4

Experiment 2. Mixed model analysis of the fixations on the object ORANGE ON TRAY *at ROI:1PP* on the tray. Intonation *is coded as* -0.5 *for NP modifier and* 0.5 *for PP modifier,* Number of Referents *is coded as* -0.5 *for 1 Referent,* 0.5 *for 2 Referents. Time (100 to 700 ms, in 100 ms intervals) is represented as an orthogonal polynomial of order two (Time^1, Time^2).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.431	0.021	20.19	<1e-04
Intonation	-0.104	0.037	-2.763	0.005
Time^1	-0.124	0.038	-3.275	0.001
Time^2	-0.122	0.023	-5.335	<1e-04
Referent	-0.036	0.036	-0.990	0.3
Intonation:Time^1	-0.126	0.031	-4.016	<1e-04
Intonation:Time^2	0.102	0.031	3.250	0.001
Intonation:Referent	0.051	0.024	2.148	0.03
Intonation:Time^1:Referent	0.138	0.063	2.196	0.02

Formula: (1 | item) + (1 | participant) + Intonation + (0 + Intonation | item) + (0 + Intonation | participant) + Time^1 + (0 + Time^1 | item) + (0 + Time^1 | participant) + Time^2 + (0 + Time^2 | item) + Referent + (0 + Referent | item) + (0 + Referent | participant) + Intonation:Time^1 + Intonation:Time^2 + Intonation:Referent + Intonation:Time^1:Referent

of Time (Time^1 and Time^2). Interestingly, however, looks increase over time in the NP modifier condition when only one referent is depicted (interaction Intonation:Referent:Time^1).

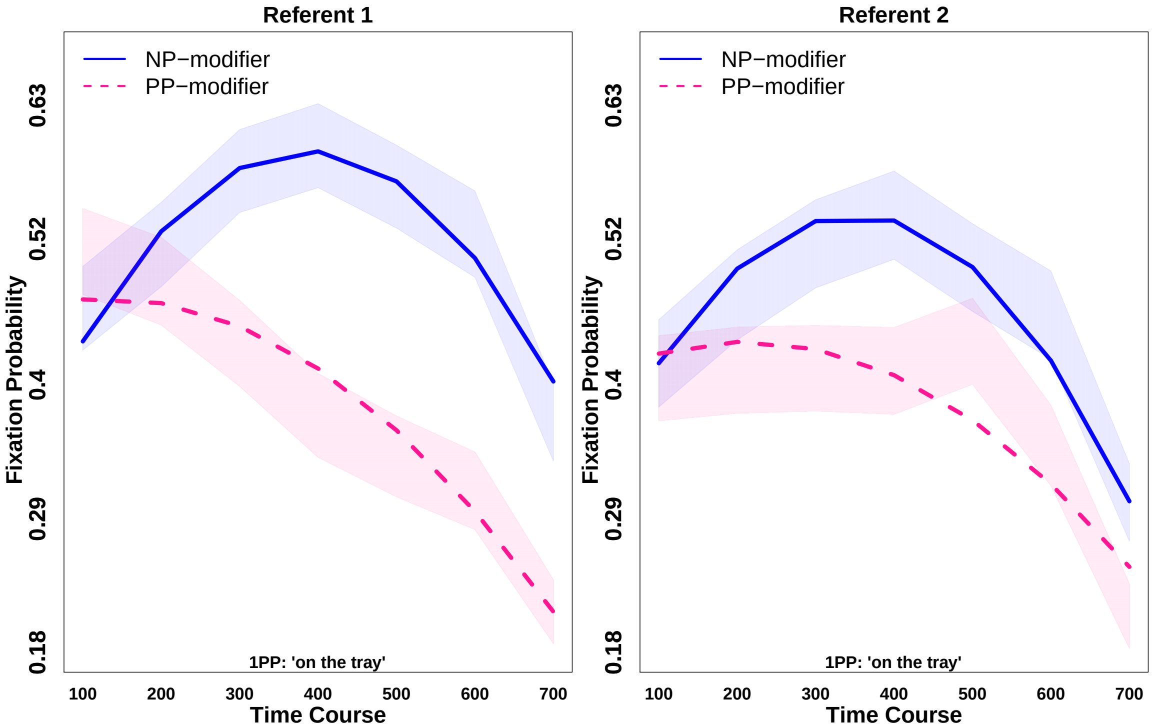


Figure 3. Experiment 2. Time course plot of fixation probability for the object ORANGE ON TRAY from 100 ms to 700 ms at ROI:1PP *on the tray*. Left panel: 1 Referent condition, right panel: 2 Referent condition. The intonation conditions (NP-modifier, PP-modifier) are marked through line types and colors. The shaded bands indicate the standard error around the observed mean. The lines represent the predicted values of the LME model reported in Table 4. Note that the offset of the region of analysis varied by items, but fixations crossing the offset were excluded, see Analysis section for details.

Experiment 3: Visual Saliency Effect on ROI:NP *the orange* on Object TRAY IN BOWL

The anticipatory effects of visual saliency are also found on the Compound Location TRAY IN BOWL, though it is reduced compared to the Single Location (see Figure 4). Table 5 shows that the effect of visual saliency is only found as an interaction with time: as the direct object *the orange* unfolds, looks on the target object increase. We also observe a significant interaction of Intonation and Time: for a PP-modifier break, we observe more looks to the Compound Location, presumably as it is interpreted as a goal location for the direct object. Crucially, we do not find a significant interaction between visual saliency and intonation, in line with the results reported for Experiment 3 in the main text.

Experiment 3: Linguistic Prominence Effect on ROI:1PP *on the tray* on Object ORANGE ON TRAY

On the object ORANGE ON TRAY for the ROI:1PP *on the tray*, we confirm the effect of intonational breaks observed in Experiment 2. In particular, an NP modifier break leads to more looks to the target object ORANGE ON TRAY, an effect that develops over time (see Figure 5). By

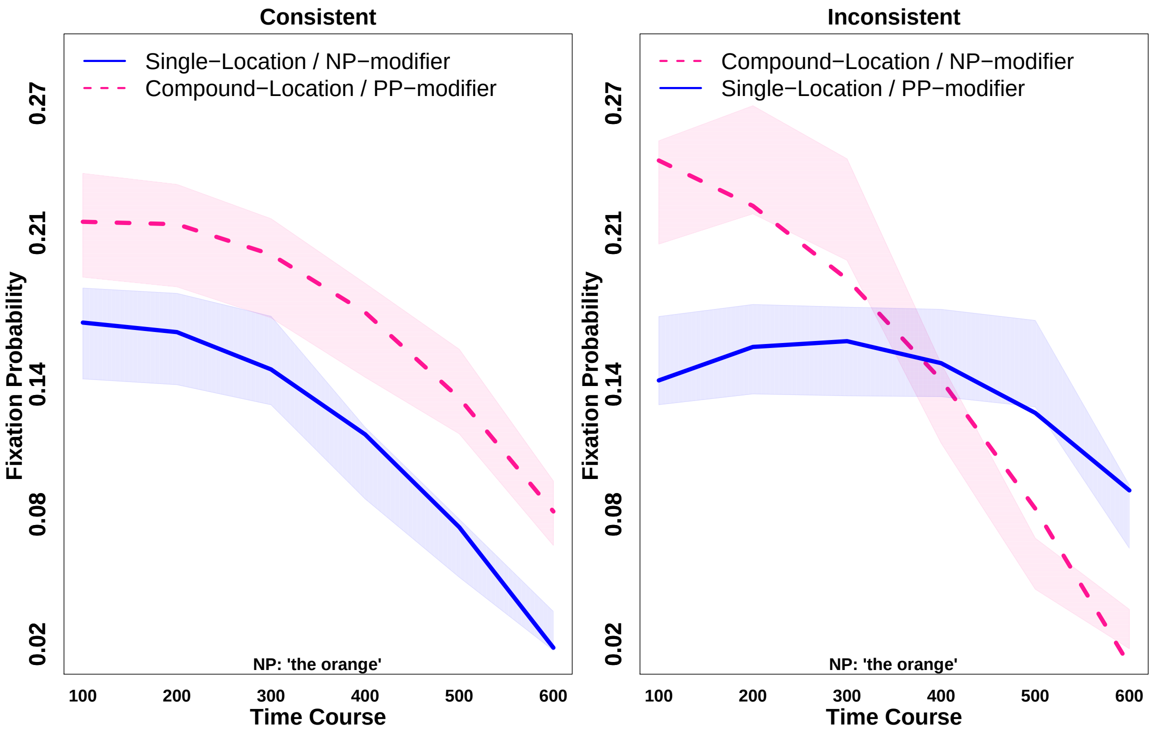


Figure 4. Experiment 3. Time course plot of fixation probability for the object TRAY IN BOWL (corresponding to the Compound Location) from 100 ms to 600 ms at ROI:1PP *the orange*. Left panel: Consistent, right panel: Inconsistent condition. The four experimental conditions are marked through line types and colors. The shaded bands indicate the standard error around the observed mean. The lines represent the predicted values of the LME model reported in Table 5. Note that the offset of the region of analysis varied by items, but fixations crossing the offset were excluded, see Analysis section for details.

definition, the effect of intonational breaks is connected with the temporal dimension of the stimulus, and therefore this effect becomes stronger as the phrase that realizes the intonation unfolds. We also find a significant interaction between Saliency and Time: when visual saliency is on the Compound Location, looks increase over time as the preposition *on the tray* unfolds. The visual saliency on the Compound Location competes for looks with the object ORANGE ON TRAY,as they both share the visual object TRAY associated to the referent *on the tray*.

Comparison Between Linear Mixed Effects Models Aggregated by Trials and by Participants

In this section, we show how a linear mixed effects model for fixation probabilities aggregated at the level of individual trials, i.e., only over temporal windows, returns coefficients comparable to those returned by a model in which the dependent measure is aggregated by participants. For this comparison, we reanalyze the fixation data of Experiment 1 on object Single Location at ROI:NP from the main text. In Table 7, we report the model coefficients obtained when fixation are aggregated at the level of individual trials (which are the results presented in the main text), whereas in Table 8, we report the coefficients obtained when the aggregation is done by participants.

Table 5

Experiment 3. Mixed model analysis of the fixations on the object TRAY IN BOWL *at ROI:NP* the orange. Intonation *is coded as* -0.5 *for NP modifier and* 0.5 *for PP modifier,* Saliency *is coded as* -0.5 *for Single Location and* 0.5 *for Compound Location. Time (100 to 600 ms, in 100 ms intervals)is represented as an orthogonal polynomial of order two (Time*^{1}*, Time*^{2}*).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.143	0.012	11.442	$<$ 1e-04
Time^1	-0.118	0.019	-6.076	$<$ 1e-04
Saliency	-0.034	0.02	-1.736	0.08
Time^2	-0.039	0.009	-4.276	$<$ 1e-04
Intonation	0.022	0.02	1.077	0.2
Time^1:Intonation	0.083	0.018	4.548	$<$ 1e-04
Time^1:Saliency	0.070	0.018	3.844	0.0001

Formula: (1 | item) + (1 | participant) + Time^1 + (0 + Time^1 | item) + (0 + Time^1 | participant) + Saliency + (0 + Saliency | item) + (0 + Saliency | participant) + Time^2 + Intonation + (0 + Intonation | participant) + (0 + Intonation | item) + Time^1:Intonation + Time^1:Saliency

Table 6

Experiment 3. Mixed model analysis of the fixations on the object ORANGE ON TRAY *at ROI:1PP* on the tray. Intonation *is coded as* -0.5 *for NP modifier and* 0.5 *for PP modifier,* Saliency *is coded as* -0.5 *for Single Location and* 0.5 *for Compound Location. Time (100 to 700 ms, in 100 ms intervals)is represented as an orthogonal polynomial of order two (Time*^{1}*, Time*^{2}*.*

Predictor	β	*SE*	*t*	*p*
Intercept	0.322	0.019	16.74	$<$ 1e-04
Time^1	-0.116	0.03	-3.812	0.0001
Time^2	-0.111	0.011	-9.977	$<$ 1e-04
Intonation	-0.075	0.031	-2.44	0.01
Saliency	0.008	0.02	0.425	0.6
Time^1:Intonation	-0.117	0.022	-5.225	$<$ 1e-04
Time^1:Saliency	0.048	0.022	2.17	0.02

Formula: (1 | item) + (1 | participant) + Time^1 + (0 + Time^1 | item) + (0 + Time^1 | participant) + Time^2 + Intonation + (0 + Intonation | participant) + (0 + Intonation | item) + Saliency + (0 + Saliency | participant) + (0 + Saliency | item) + Time^1:Intonation + Time^1:Saliency

By comparing the two tables can be seen that all factors of the model based on aggregation by trials also figure in the model based on aggregation by participants. The estimates of the coefficients obtained with the different types of aggregation are close to each other: for SaliencySingleLoc, for example, we obtain $\beta = 0.094$ for aggregation by trials and $\beta = 0.098$ for aggregation by participants. A linear-mixed effect model with both participant and items both as random effects is known to return a smaller Type 1 error rate (Baayen, Davidson, & Bates, 2008). This comparison therefore suggests a use of mixed effect models containing random intercept and slopes for both participants and items, i.e., using by trials models with aggregation on temporal windows only.

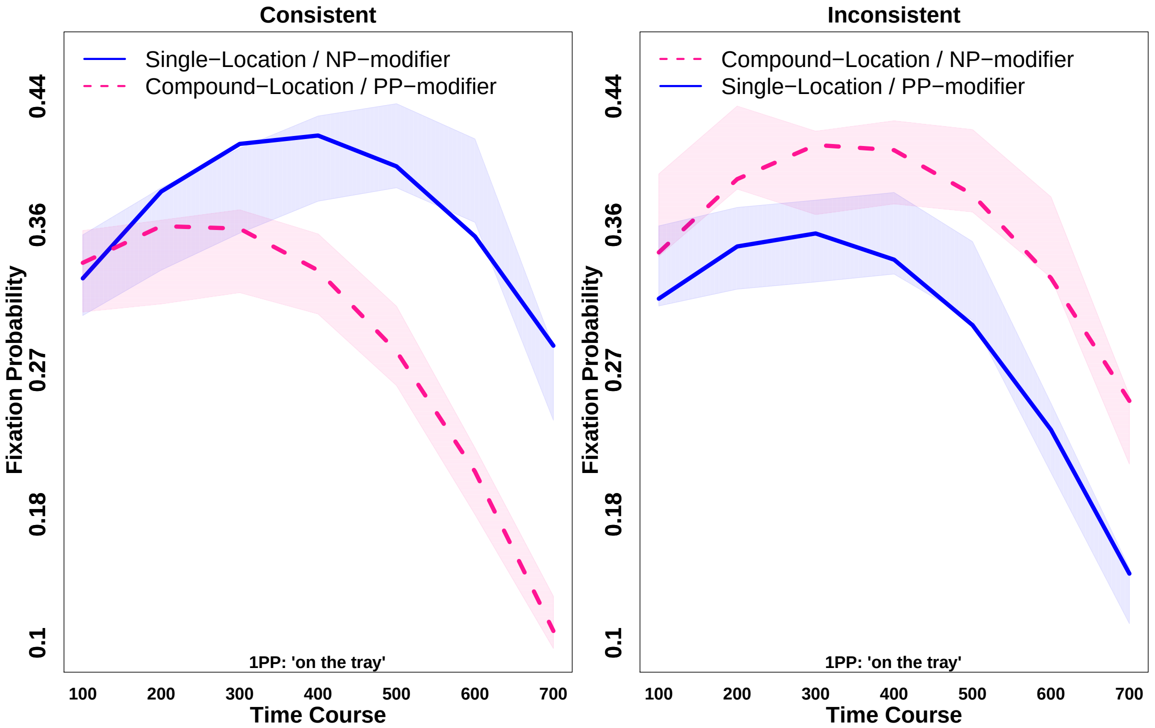


Figure 5. Experiment 3. Time course plot of fixation probability for the object ORANGE ON TRAY from 100 ms to 700 ms at ROI:1PP *on the tray*. Left panel: 1 Referent condition, right panel: 2 Referent condition. The intonation conditions (NP-modifier, PP-modifier) are marked through line types and colors. The shaded bands indicate the standard error around the observed mean. The lines represent the predicted values of the LME model reported in Table 6. Note that the offset of the region of analysis varied by items, but fixations crossing the offset were excluded, see Analysis section for details.

Time-course Plots across the Whole Sentence for Single-Location, Compound-Location and Other objects.

In this section, we provide time-course fixation probability plots of the objects (Single-Location, Compound-Location, and Other objects except the Background) across the four experimental conditions of Experiment 3. We choose to visualize Experiment 3, as we manipulated both visual saliency and linguistic prominence in this experiment.

As explained in the main text, individual sentences vary in the length of the phrases they contain. Therefore, the whole sentence plots presented here can only be descriptive, they cannot be used to statistically assess whether there are significant difference in the fixation patterns at specific linguistic ROI. For this, the ROI need to aligned based on their onset, which is what we have done for all analyses presented in the main text.

In spite of this, some effects are discernible even in the whole sentence plots. In Figure 6(a), for example, where visual saliency is on the object Single Location, it can be clearly seen that fixation probability on the salient object steeply increases after the onset of the scene, and then rapidly decays prior to the onset of the sentence (a more thorough analysis of this phenomena can

Table 7

Aggregation by trials: Mixed model analysis of the fixation probability on the object BOWL *(corresponding to the Single Location) at ROI:NP* the orange. Saliency *is contrast coded with* No Saliency *as reference level,* Number of Referents *is coded as* -0.5 *for 1 Referent, 0.5 for 2 Referents. Time (100 to 600 ms, in 100 ms intervals) is represented as an orthogonal polynomial of order two (Time*^{1}*, Time*^{2}*).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.106	0.011	8.966	<1e-04
Time^{1}	-0.07	0.017	-4.067	<1e-04
SaliencySingleLoc	0.094	0.022	4.191	<1e-04
Time^{2}	-0.041	0.008	-4.751	<1e-04
Referent	-0.026	0.017	-1.501	0.1
SaliencyCompoundLoc	-0.051	0.018	-2.786	0.005
SaliencySingleLoc:Time^{1}	-0.089	0.021	-4.260	<1e-04
SaliencySingleLoc:Referent	0.053	0.017	3.092	.002
Referent:Time^{1}	-0.044	0.017	-2.548	.01

Formula: (1 | item) + (1 | participant) + Time^{1} + (0 + Time^{1} | item) + SaliencySingleLoc + (0 + SaliencySingleLoc | item) + (0 + SaliencySingleLoc | participant) + Time^{2} + Referent + (0 + Referent | participant) + (0 + Referent | item) + SaliencyCompoundLoc + (0 + SaliencyCompoundLoc | item) + (0 + SaliencyCompoundLoc | participant) + Time^{1}:SaliencySingleLoc + SaliencySingleLoc:Referent + Time^{1}:Referent

be found in section). This indicates that any saliency effects observed after speech onset can be attributed to an interaction between saliency and the linguistic information being processed, rather than being an effect of saliency alone (such effects seem to occur only at the onset of the scene)

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506.

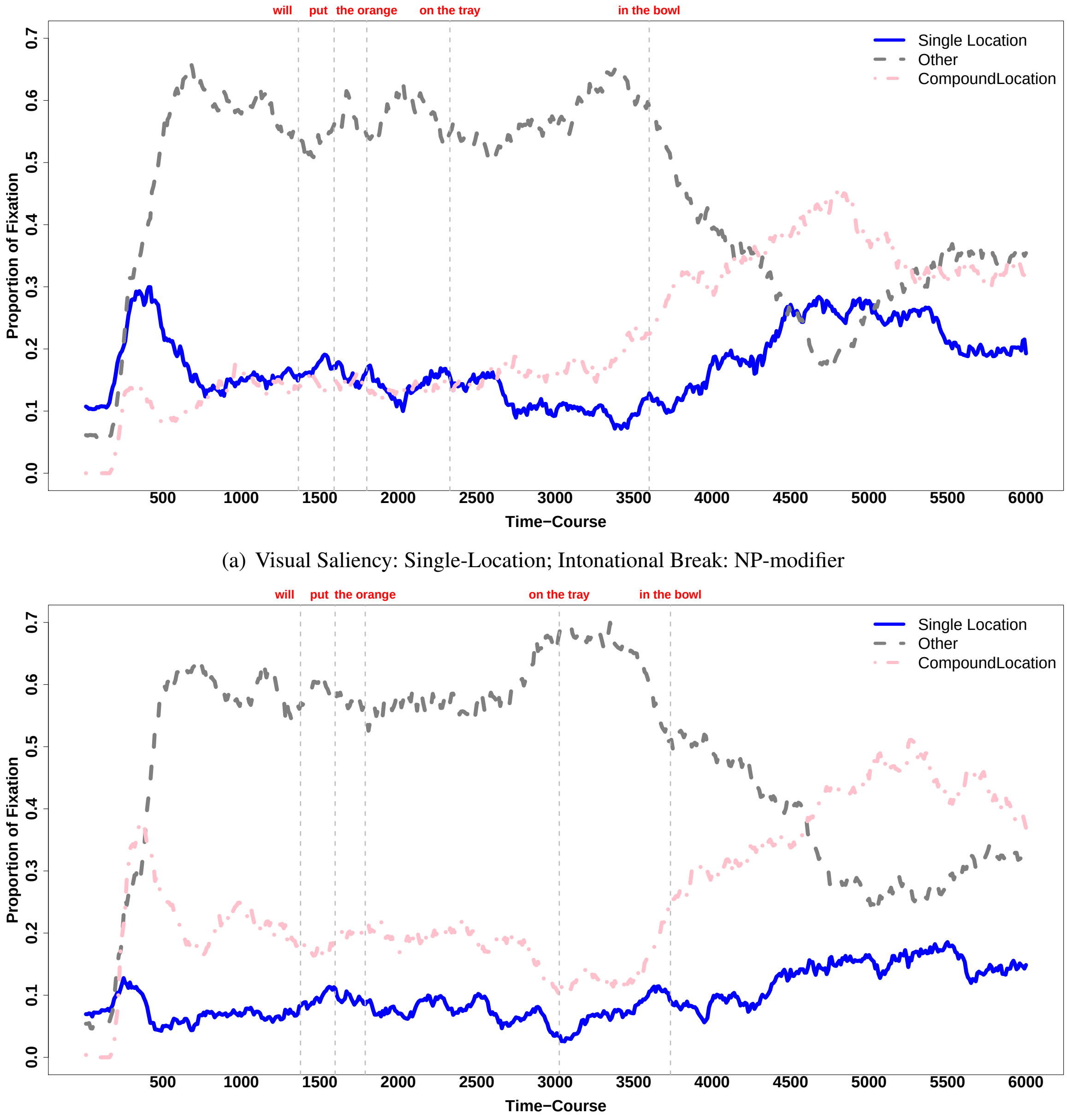
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*(268), 632–634.

Table 8

Aggregation by participants: Mixed model analysis of the fixation probability on the object BOWL *(corresponding to the Single Location) at ROI:NP* the orange. Saliency *is contrast coded with* No Saliency *as reference level,* Number of Referents *is coded as* -0.5 *for 1 Referent, 0.5 for 2 Referents. Time (100 to 600 ms, in 100 ms intervals) is represented as an orthogonal polynomial of order two (Time*^{1}*, Time*^{2}*).*

Predictor	β	*SE*	*t*	*p*
Intercept	0.1	0.006	14.532	<1e-04
SaliencySingleLoc	0.098	0.016	5.791	<1e-04
Time^1	-0.067	0.007	-9.333	<1e-04
Time^2	-0.039	0.007	-5.428	<1e-04
Referent	-0.028	0.012	-2.239	0.02
SaliencyCompoundLoc	-0.051	0.013	-3.868	0.0001
SaliencySingleLoc:Time^1	-0.107	0.02	-5.291	<1e-04
Referent:SaliencyCompoundLoc	0.055	0.014	3.85	0.0001
Time^1:Referent	-0.04	0.014	-2.791	0.005
Time^1:SaliencyCompoundLoc	0.048	0.02	2.375	0.01
SaliencySingleLoc:Time^1:Referent	-0.081	0.035	-2.312	0.02

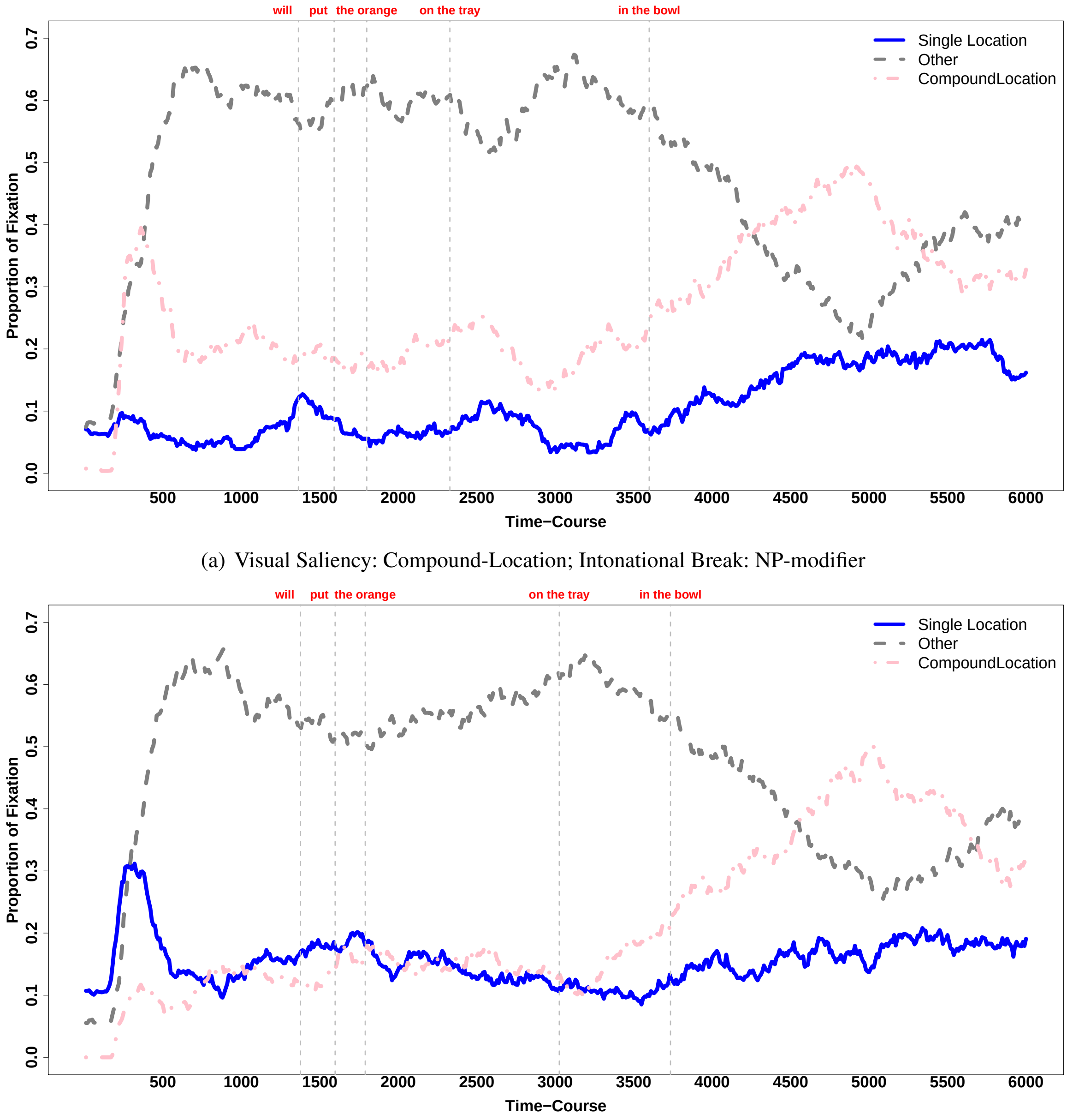
Formula: (1 | participant) + SaliencySingleLoc + (0 + SaliencySingleLoc | participant) + Time^1 + Time^2 + Referent + (0 + Referent | participant) + SaliencyCompoundLoc + (0 + SaliencyCompoundLoc | participant) + SaliencySingleLoc:Time^1 + Referent:SaliencyCompoundLoc + Time^1:Referent + Time^1:SaliencyCompoundLoc + SaliencySingleLoc:Time^1:Referent + Time^1:Referent



(a) Visual Saliency: Single-Location; Intonational Break: NP-modifier

(b) Visual Saliency: Compound-Location; Intonational Break: PP-modifier

Figure 6. Time-course fixation probability over the entire sentence across the four experimental condition of Experiment 3 for all objects (part 1).



(a) Visual Saliency: Compound-Location; Intonational Break: NP-modifier

(b) Visual Saliency: Single-Location; Intonational Break: PP-modifier

Figure 7. Time-course fixation probability over the entire sentence across the four experimental condition of Experiment 3 for all objects (part 2).