Temporal Dynamics of Scan Patterns in Comprehension and Production

Moreno I. Coco (M.I.Coco@sms.ed.ac.uk) and
Frank Keller (keller@inf.ed.ac.uk)
Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, UK

Abstract

Speakers and listeners in a dialogue establish mutual understanding by coordinating their linguistic responses. When a visual scene is present, scan patterns on that scene are also coordinated. However, it is an open question which linguistic and scene factors affect coordination. In this paper, we investigate the coordination of scan patterns during the comprehension and generation of scene descriptions. We manipulate the animacy of the subject and the number of visual referents associated with it. By using Cross Recurrence Analysis, we demonstrate that coordination emerges only during linguistic processing, and that it is especially pronounced for inanimate unambiguous subjects. When the subject is referentially ambiguous (more than one visual object associated with it), scan pattern variability increases to the extent that the animacy effect is neutralized.

Keywords: Scan patterns, situated language processing, cognitive dynamics, coordination

Introduction

When language is comprehended or produced in a visual context, information about fixated objects has to be integrated with the linguistic information that is concurrently processed (e.g., Spivey-Knowlton et al. 2002); this integration requires visual attention and sentence processing to be synchronized temporally (e.g., Zelinsky & Murphy 2000). Language comprehension and language production, however, differ in their temporal interaction with visual attention. In a comprehension task, visual attention is guided by linguistic information, and its main role is to anticipate which objects the speech could refer to next (e.g., Altmann & Kamide 1999). In a production task, instead, visual attention plays an active role in deciding which objects in the scene should be mentioned in a sentence (e.g., Griffin & Bock 2000).

The relation between comprehension and production has been investigated mainly in the context of dialogue. A prominent account of how comprehension and production relate to each other is the interactive alignment model (Pickering & Garrod, 2007); which assumes that successful dialogue leads to aligned representations at every linguistic level, and that this alignment is supported by priming, i.e., the reuse of linguistic material.

Importantly, this process of alignment in dialogue has been observed to go beyond aligned linguistic representations; it also includes the gaze coordination of dialogue partners. Richardson et al. (2007) showed that the scan patterns of listeners and speakers engaged in a dialogue about six characters are coordinated. This coordination is subject to a characteristic temporal lag, with the same character being fixated consistently later by listeners than by speakers. This confirms that visual responses during comprehension are launched after the linguistic material is understood; whereas in production, visual responses are launched prior or during sentence generation.

These results strongly suggest the existence of alignment mechanisms that underlie the coordination of comprehension and production processes. However, especially with respect to the evidence for gaze coordination, it is unclear what the role og visual and linguistic information is, and whether the characteristic lag underlying gaze coordination depends on such information.

In Richardson et al. (2007), in fact, the visual information available to the participants is not naturalistically situated (i.e., six portait pictures of characters from TV serials), and the linguistic information used by the speaker to guide the listener, besides referring to a depicted character, does not actively interact with it. As a result of this, the gaze coordination obtained in the dialogue is achieved through a shallow process of character identification: the speaker is talking about X and the listener looks at X with a constant delay.

In this paper, we present a study in which we explicitly investigate how linguistic and visual information interact to produce coordinated scan patterns. We explore coordination at different levels of granularity, from the macro-level of the whole trial down to the level of individual objects. Moreover, we test how coordination is influenced by the visual and linguistic referential information shared in comprehension and production, focusing on the animacy of the subject of the sentence, shown to influence both linguistic and visual responses (Coco & Keller, 2010), and the number of targets (visual referents associated with the subject).

Our main hypothesis is that the characteristic lag underlying the scan pattern coordination between comprehension and production emerges only when sentence processing is actively involved, and that it is directly influenced by the properties of the visual and linguistic information being processed. In particular, scan patterns are expected to show less coordination on a single animate target, as the associated information spans a wider range of contextual possibilities. In contrast, the low linguistic relevance of an inanimate target, and the referential ambiguity of multiple targets should force participants to depend more strongly on contextual scene information, thus triggering a higher degree of coordination.

Experiment

Our study aims to explore the role of referential factors in the temporal dynamics of scan pattern coordination between language comprehension and production during the description

In Laura Carlson, Christoph Hoelscher, and Thomas F. Shipley, eds., *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 2302–2307. Boston, 2011.

of naturalistic scenes.

Processing descriptions requires visual and linguistic referential information to be overtly integrated. When a description is produced, active processes of scene exploration interact with the encoding of linguistic information; when such a description is instead understood, its decoding is constrained and modulated by the visual information available.

The main goal of the current study is to test whether the temporal dynamics of scan pattern coordination between comprehension and production of scene descriptions differs from that observed in dialogue. Additionally, we test whether referential factors pertaining to the linguistic and visual information processed modulate the associated pattern of coordinated gazes.

Method

We quantify coordination by using eye-tracking data collected in the two independent experiments (production and comprehension), which involve the same visual and linguistic stimuli.

In an eye-tracking language production experiment (Coco & Keller, 2010), we asked participants to describe a photo-realistic scene after being prompted with a target word, which was either animate or inanimate (e.g., *man* or *hat*), and corresponded to either one, two or three visual objects depicted in the scene. The production data considerably varies in sentence and scan pattern complexity. Thus, in order to control this variability and have sentences with similar syntactic structures and controlled semantic factors, we select a subset of 24 sentences (together with the associated scan patterns), produced by different participants, to be used in the follow-up language comprehension experiment.

We followed three criteria to select this subset: (1) the sentence is transitive and mentions only two visual referents, e.g., *the* WOMAN *is playing the* VIOLIN, making it possible to test coordination on a precise number of individual objects (kept constant across the set), (2) the subject of the sentence is either animate or inanimate, which allows us to observe how the conceptual property of animacy modulates coordination, and (3) the target object associated with the subject is either unique (i.e., there is one corresponding visual object) or referentially ambiguous (i.e., there are three corresponding visual objects), to assess the role of ambiguity resolution. Figure 1 depicts a set of example stimuli. The 24 sentences we selected represent a design with four conditions (six sentences per condition), crossing the factors Animacy (animate or inanimate) and Number of Targets (one or three). These sentences were played to a different set of participants in an indepedent language comprehension experiment while they viewed the associated scenes. For this purpose, the sentences were recorded by a female native speaker of English.

Figure 1: Example of experimental conditions and materials (scenes and sentences).

Procedure Forty-eight (24 per task) native speakers of English, all students of the University of Edinburgh, were each paid five pounds for taking part in the experiment. An EyeLink II head-mounted eye-tracker was used to monitor participants' eye-movements with a sampling rate of 500 Hz. Images were presented on a 21" multiscan monitor at a resolution of 1024 x 768 pixels. Participants sat 60–70 cm from the computer screen, which subtend a region of approximately 20 degrees of visual angle. Only the dominant eye was tracked. In the description task, a target word appeared for 750 ms at the center of the screen, after which the scene followed. A lapel microphone was used to record the descriptions generated. In the comprehension task, participants had a scene preview of 1500 ms before the sentence was played.

A nine points randomized calibration was carried out at the beginning of each experiment, and repeated approximately every 24 trials. Drift correction was performed before each trial. Once every four trials, during the comprehension task, a yes/no comprehension question about the content of the scene or the sentence was asked. Participants had to respond by pressing a button on a control pad. In the description task, there was no time limit for the trial, and to pass to the next trial, participants pressed a button on the response pad. In the comprehension task, the trial ended 1500 ms after the end of the sentence. Both experimental tasks were explained using written instructions and took about 30 minutes to complete.

Analysis

The temporal dynamics governing the interaction between visual attention and language processing are different for comprehension and production. In comprehension, visual responses are linked to sentence processing only when the sentence is listened to; in production, instead, visual responses interact with sentence processing both prior and during the mention of a visual object.

A way to investigate temporal variability between two time-series while exploring the underlying regularity is Cross Recurrence Analysis (CRA, Marwan & Kurths 2002; Richardson et al. 2007).

Nominal Cross Recurrence Analysis Conceptually, CRA compares two time series by calculating the degree of their recurrence when **delays** are introduced with different levels of **phase space** embedding. From an original time-series $X(t)$,

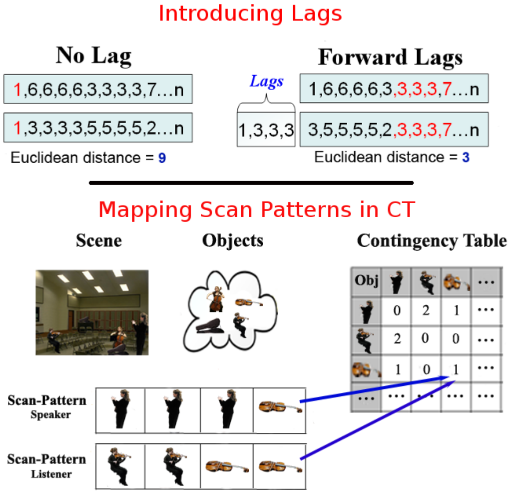

Figure 2: The top of the figure shows a simplified example of how lags are introduced in the time series and cross-recurrence calculated. The bottom part shows how a contingency table mapping the information of object co-occurrences between two scan patterns is created.

delayed copies $X(t+\tau)$ are generated by introduction a lag τ into the original time series. The different dimensions of phase space embedding are obtained by considering multiple lags $X(t+m\tau)$. The lag is introduced to compare one time series with the future or the past of itself, or to compare it to another time series. The phase space consists of the different intervals over which the delays are assigned. Over these time-delayed copies and across the different lags, **recurrence,** i.e., a measure of similarity, is calculated.

Suppose we have scan-pattern data from two participants, each represented as a sequence of numbers (see Figure 2). Participant 1 was producing a sentence and participant 2 was listening to it. If the two sequences are not shifted and we measure their similarity by computing, for example, Euclidean distance, we obtain a distance of 9. If we shift the time series of participant 2 by moving his sequence forward by four time units, we observe increased similarity: the Euclidean distance is now 3. The interpretation is that more time was needed by participant 2 to produce a sequence similar to participant 1. Since our time series are scan patterns, i.e., sequences of fixated objects, we follow Dale et al. (2011) and adopt a categorical version of CRA, where recurrence is obtained by means of contingency tables; refer to Figure 2 for an example. At each lag τ, we construct a contingency table CT, which is a square matrix with the objects of a given scene as its rows and columns. Each element of this matrix represents the number of times the pair of objects (i,j) co-occurs between the two scan patterns x and y. More formally: $CT_{i,j}(\tau) = \sum_{t=1}^{t=T-\tau} q(t)$, where T is the length of each scan pattern and $q(t) = 1$ if $x(t) = i$ and $y(t+\tau) = j$, and $q(t) = 0$ otherwise.

From CT, we can compute two measures of recurrence: matching recurrence RR and object-specific recurrence ϕ_k. Matching recurrence is computed along the diagonal of CT by adding the frequencies of looks to the same objects. Often, however, we are interested in the agreement between the two scan patterns on a specific object k. This information is obtained by computing the ϕ_k coefficient, which increases with the frequency of matching looks on the same object $(k-k)$ and away from this object $(\neg k - \neg k)$ between the two scan patterns. On the other hand, ϕ_k decreases with the frequency of mismatching objects $(k - \neg k$, and vice versa); refer to Dale et al. (2011), for more details.

Regions of Analysis In order to capture how the temporal dynamics of coordination is influenced by the introduction of linguistic information, we conduct our analysis at three levels: global, phase, and object.

In the global analysis, similar to Richardson et al. (2007), we look at the whole trial. At this macro-level, we observe how recurrence develops across different lags and measure the impact of subject animacy and visual referential ambiguity on recurrence. If these two factors do not influence coordination, a similar amount of recurrence should result in all conditions.

In the phase analysis, we compute recurrence separately **before** and **during** sentence processing, and explore the distribution of optimal lags (i.e., the lags associated with maximal recurrence for each pair of scan patterns) associated with the visual objects describing the subject and object of the sentence. Before sentence processing starts, we do not expect any specific temporal correlation between comprehension and production, as visual attention is not yet guided by linguistic information to the same target objects. However, during production, in line with previous literature (Richardson et al., 2007), we expect recurrence to increase when the scan patterns of production are delayed with respect to the scan patterns of comprehension, i.e., when a positive lag is introduced.

In the object analysis, we evaluate how recurrence (measured as ϕ_k) changes for the visual objects associated with the linguistic referents of the sentence (subject and object), before and during sentence processing, across the different conditions of Animacy and Number of Targets. Before production, we expect higher recurrence on the second object, as it usually represents the receiver of an action (for an animate subject), or an object spatially related to the subject (if the subject is inanimate). We hypothesize that in preparation for sentence processing, both in comprehension and in production, visual attention explores the different possible events taking place in the scene that could be referred to linguistically. As animate actors are quickly spotted, visual attention focuses more on the receiver of the action. Similarly, in the case of an inanimate subject, attention must be directed to other inanimate objects that could be spatially related to the subject.

Inferential Analysis We use linear mixed effect models (Baayen et al., 2008) to quantify the impact of Lags, Animacy, and Number of Targets on the amount of recurrence observed. A linear mixed effect model is a multilevel extension of linear regression, where the regression coefficients of explanatory variables (fixed effects) on a dependent measure are inferred with respect to random effects, usually related to sampling variables, such as participants.

We use and report estimates of LME coefficients for the global and object analysis, where the dependent measures are recurrence and ϕ_k. Our predictors are Lag, Number of Targets (one, three), Animacy (animate, inanimate), and for the object analysis, we also include Object (first, second), a categorical variable indicating the visual objects referred to in the sentence.[1] The random effects are Participants, both in comprehension and production, and Scenes.

Since our explanatory variables can be influenced by the variability of scene configurations, we residualize our dependent measures prior to the LME analysis, against three variables (Clutter, Referents, and Area) related to each individual scene. Clutter quantifies the visual density of the scene (Rosenholtz et al., 2007), Referents describe the total number of visual objects in a scene, and Area is the number of pixels occupied by the visual objects associated with the linguistic referents of the sentence.

All fixed factors were centered to reduce collinearity. The mixed models were built following a forward step-wise procedure. We start with an empty model, then we add the random effects. Once all random effects have been evaluated, we proceed by adding the predictors. They are added one at time and ordered by log-likelihood improvement of model fit; the predictor that improves most model fit is added first. Every time we add a new parameter to the model (fixed or random), we compare its log-likelihood against the previous model. We retain the additional predictor if log-likelihood fit improves significantly ($p < 0.05$). The final model is therefore the one that maximizes model fit with the minimal number of predictors.

Results and Discussion

We present three analyses: (1) in the global analysis, we explore how recurrence changes with lags for the whole trial in the different conditions; (2) in the phase analysis, we investigate changes in recurrence by examining before and during sentence processing. We search for the lag maximizing recurrence, and observe whether it differs for the two visual objects of the sentence; (3) in the object analysis, we explore how the experimental factors Animacy, Number of Targets, and Object interact with recurrence before and during sentence processing.

Global: Recurrence for the Whole Trial In Figure 3, we show mean and confidence intervals of recurrence calculated on scan patterns generated during the whole trial across different lags with maximum lag ±3500 ms. We observe differences in the magnitude and trend of recurrence across conditions.

In particular, sentences with inanimate subjects trigger a higher scan pattern recurrence compared to sentences with animate ones ($\beta_{Inanimate} = 0.036; p = 0.07$); and this effect reaches significance when there is only one visual target ($\beta_{Inanimate:Target-One} = 0.3; p < 0.05$)

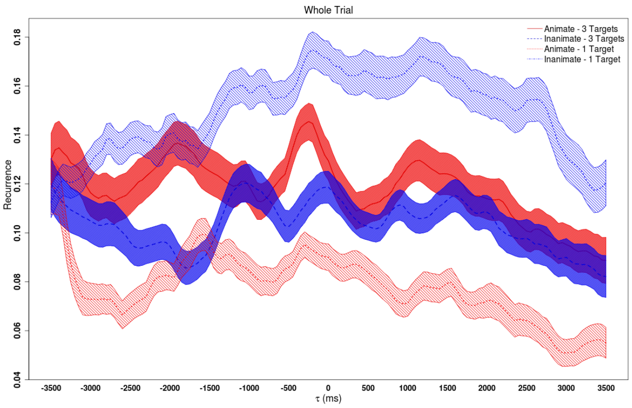

Figure 3: CRA: Mean and confidence intervals of scan pattern recurrence during the whole trial for the different lags ($\tau = \pm 50$ ms from -3500 ms to 3500 ms). Line density indicates the number of targets (three: high density, one: low density), color indicates the animacy of the subject (red: animate, blue: inanimate).

Animate objects are linked to a larger set of event relations within a given scene compared to inanimate objects, which instead are often contextualized by their spatial relation with another object. When interpreting the coordination of gazes between comprehension and production, this implies that for an inanimate single target, once sentence processing starts and the subject is spelled out, it is much easier to guess which object is going to be mentioned next. In contrast, the competition generated by the visual ambiguity in the Three Targets condition tends to increase variability of scan patterns, making responses in the animate and inanimate condition more similar. Nevertheless, three animate referents attract more visual attention than three inanimate referents, especially when linguistic information is not yet introduced; which explains the positive interaction between Animate subject and Three Targets ($\beta_{Animate:Target-Three} = 0.3; p < 0.05$).

It is important to note that at the global level of analysis, we fail to find an effect of lag ($\beta_{Lag} = -0.00005; p > 0.1$). This differs from the findings of Richardson et al.'s (2007) study, which is based on trials that consist of dialogues. In a dialogue, the speaker provides the listener with linguistic guidance throughout the whole trial; whereas in descriptions[2], the linguistic guidance to listeners (expected to improve gaze coordination) is limited to when the description is actually mentioned. Thus, we expect that the characteristic lag observed by Richardson et al. (2007) should emerge only during sentence processing. To test this, we analyze what happens before and during sentence processing separately.

Phase: Lag Distribution Before and During Processing In Figure 4, we plot the frequency distribution of the lags

---

[1] If the sentence is *the woman is playing the violin*, then Object: first is the recurrence on WOMAN-L (we disambiguate multiple visual referent by their position in the scene), whereas the Object: second is the recurrence on VIOLIN.

[2] Notice, our speakers and listeners do not interact, as they are tested in two independent experiments.

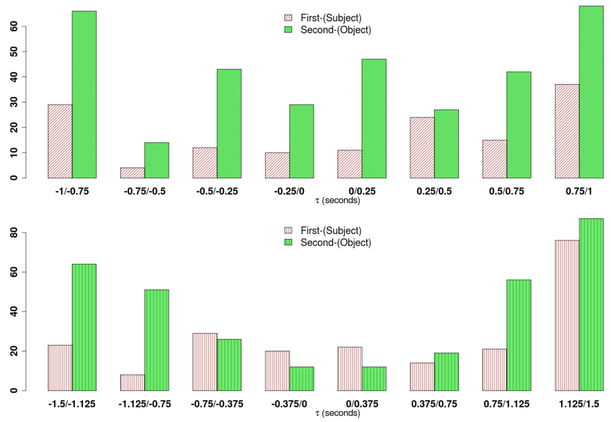

Figure 4: Frequency distribution of optimal lags: before (top panel) and during (bottom panel) sentence processing. The optimal lag is the one that gives maximal scan pattern recurrence on the visual referents associated to the sentence (first, i.e., subject; second, i.e., object).

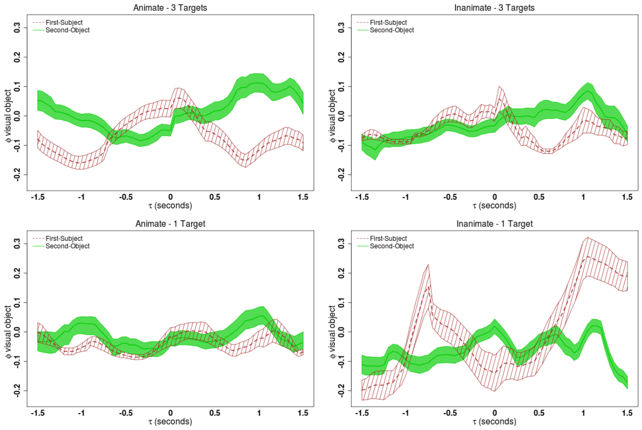

Figure 5: ϕ_k coefficient of visual objects associated with the sentence (first, i.e., subject; second, i.e., object), during sentence processing.

which give maximal recurrence, before and during sentence processing, on the visual referents associated with the sentence.

Before sentence processing, the introduction of lags improves recurrence on both visual objects; nevertheless, this increase in recurrence does not relate to any specific direction of temporal shifts. This implies that at lag zero, production and comprehension have highly dissimilar scan patterns, but they tend to be more aligned when delays are introduced. Probably, visual attention tends to converge on a similar set of objects when a certain time has elapsed in both processes of comprehension and production. When looking at the objects, we find maximal recurrence more often in relation to the second visual object (the sentence object). We argue that visual attention focuses more on the objects in the scene, which are either receivers of actions, or are in a spatial relation to other objects, as they carry important causal information to understand the event taking place in the scene.

During sentence processing, we observe a clear trend of maximal recurrence for positive lags. In line with previous literature (Richardson et al., 2007), a scan pattern generated during description needs to be shifted forward with respect to the associated comprehension scan pattern, as visual referents are usually fixated before description in production, but identified in comprehension after the associated linguistic referents have been listened to. It is important to notice how during sentence processing, recurrence on the first visual object (the sentence subject) increases substantially already at lag zero compared to before sentence processing. Naturally, since the sentence starts with the subject, visual attention is oriented immediately to the associated visual referent. Moreover, we find that increasing positive lags improve recurrence on this visual referent. A similar increase is seen also for the second object, but it holds for both positive and negative lags. Furthermore, in general, it is clear that the relative gain in recurrence by shifting is higher during sentence processing than before, for both objects. This points to the important role played by lags in aligning comprehension and production during the activation of sentence processing mechanisms.

Object: Influence of Animacy and Number of Targets In Figure 5, we show how the ϕ_k of the first and second object changes across conditions, during sentence processing[3].

We observe a main effect of Lag, where the ϕ_k coefficient of both objects gains by a positive shift of production with respect to comprehension ($\beta_{Lag} = 0.0011; p < 0.05$) This confirms that during sentence processing itself, the coordination of scan patterns in comprehension and production occurs with a characteristic delay (Richardson et al., 2007). Moreover, we find an interaction between Lag and Object, such that the second object gains more by positive shifting ($\beta_{Lag:Object-Second} = -0.0008; p < 0.0001$). Once the subject of the sentence, i.e., the first object, is identified, visual attention focuses on the second object, which is the receiver of the action, for animate subject, or on a spatially related object, for inanimate subjects. It is also interesting to note that when the subject of the sentence is associated with a single inanimate target, we observe substantial gains when shifting on the corresponding object, i.e., the first object, in both temporal directions, with the highest peak found at positive lags. In order to understand this result, it is important to remember that the ϕ_k coefficient penalizes mismatches. So, ϕ_k is positive when gazes are either both on the target object (e.g. violin,violin) or both on completely different objects (e.g. woman-L, woman-R); ϕ_k is instead negative when there is a mismatch, i.e. one gaze on the target object, the other on a different object (e.g. violin,woman-L).

An inanimate object has a low linguistic relevance, hence there is a high chance that is unattended if visual attention is not directed towards it by a cue. So, the first positive peak observed at negative lags indicates that when the alignment between comprehension and production widens, gazes tends to be on completely different objects. However, at positive

---

[3] We focus on the during sentence processing phase, as the LME analysis in the before analysis failed to yield any significant results.

lags, we observe a second and highest peak, which probably reflects gaze agreement on the inanimate target object. In fact, once the inanimate object has been mentioned, visual attention needs to locate and retrieve information about it. This process generates a delay in comprehension, which is reflected by the larger gain in recurrence when the production scan patterns are shifted forward.

General Discussion

The processes of language comprehension and production share cognitive mechanisms which are intimately connected. Research in dialogue has shown coordination between speakers and listeners both in their linguistic and visual responses (Pickering & Garrod, 2007; Richardson et al., 2007). However, previous work fails to identify the factors involved in the coordination of visual responses in production and comprehension, and the temporal dynamics underlying it.

In this paper, we investigated the temporal aspects of scan pattern coordination during the generation and comprehension of scene descriptions. Descriptions, in contrast to dialogues, allow us to pin down more precisely the influence of shared referential information during the overt interaction between visual and linguistic responses.

In order to quantify the temporal dynamics underlying gaze coordination, we used Cross Recurrence Analysis: a technique used to unravel recurring patterns between time series (Marwan & Kurths, 2002; Dale et al., 2011). In line with Richardson et al. (2007), we found substantial recurrence between scan patterns in production and comprehension, but we also observed important differences across phases of analysis and across individual objects. These differences are modulated by both the animacy of subject and the number of targets. In particular, we find that delays in production increase coordination with comprehension during sentence processing (but not before), and also improve the agreement between scan patterns on the visual object identifying the subject of the sentence. We argue that prior to the availability of linguistic information, visual attention focuses on objects which are either receivers of actions, or objects that are involved in a spatial relation with the target; the subject is in focus only if it is explicitly mentioned.

The number of targets corresponding to a certain object interacts with their animacy in several interesting ways. A single animate object in the scene generates more variability between the scan patterns, which manifests itself in the recurrence remaining zero at all lags. This is perhaps due to the larger amount of conceptual knowledge related to animate objects, which offers participants a wider space of contextual relations within the scene. An inanimate single object, on the other hand, has a more limited contextual potential; therefore coordination between scan patterns in production and comprehension becomes easier, and is attained for positive lags. When multiple visual objects are associated with the same subject referent, the influence of its animacy is neutralized, due to the ambiguity introduced.

In future work, we are planning to address some shortcomings of the study presented here. In particular, we are planning a cooperative version of the description task, in which speakers and listeners are simultaneously recorded and asked to interact. The co-presence of speaker and listener allow us to have a more controlled and counterbalanced design both in terms of experimental conditions and data accuracy (e.g., equal numbers of speakers and listeners). Moreover, a cooperative task give us the possibility to explore how the interaction of different cognitive processes, e.g., motor actions and visual responses, modulates the cross-modal coordination between comprehension and production.

Overall, we have shown that scan pattern coordination is a key mechanism that enables the integration of comprehension and production processes. Crucially, we demonstrated that there are important visual and linguistic factors which need to be accounted for in order to achieve a full understanding of the cognitive dynamics underlying this integration.

Acknowledgments

The support of the European Research Council under award number 203427 ”Synchronous Linguistic and Visual Processing” is gratefully acknowledged.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

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.

Coco, M., & Keller, F. (2010). Sentence production in naturalistic scene with referential ambiguity. In *R. Catrambone & S. Ohlsson (Eds.), Proceedings of the 32th Annual Conference of the Cognitive Science Society, Portland*.

Dale, R., Warlaumont, A., & Richardson, D. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos.*.

Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*, 274–279.

Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, *302*, 299–307.

Pickering, M., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*, 105–110.

Richardson, D., Dale, R., & Kirkham, N. (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological science*, *18*, 407–413.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*, 1–22.

Spivey-Knowlton, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of syntactic context on syntactic ambiguity resolution. *Cognitive Psychology*, (pp. 447–481).

Zelinsky, G., J., & Murphy, G. (2000). Synchronizing visual and language processing. *Psychological science*, *11*, 125–131.