Do embodied effects reflect the consequences rather than the causes of meaning? - "Ja!"

Fabian Tomaschek[*1], Michael Ramscar[1], Samuel Thiele[2], Barbara Kaup[2], Harald Baayen[1]

[1]Department of Linguistics, University of Tübingen, 72074 Tübingen, Germany
[1]Department of Psychology, University of Tübingen, 72074 Tübingen, Germany

* Corresponding Author (fabian.tomaschek@uni-tuebingen.de), Department of Linguistics, University of Tübingen, 72076 Tübingen, Germany

Abstract

It has been shown that a movement in a direction incongruent with the spatial semantics of words typically requires more time than movements that are directionally congruent. Two explanations have been proposed for this effect. Either a word's meaning is understood by using an internal model to simulate a word's meaning – and incogruent directionality needs time to be resolved. Or words simply serve to reduce hearers' uncertainty about future states of the world, facilitating actions that prepare for them. However, since previous experiments have focused on actions that are directly involved in the exploration of space, they provide evidence for both hypotheses. Experiment 1 of the present study avoids this shortcoming. We investigated the basic downwards directed articulatory gesture producing a high-frequency German word, *ja* ('yes'), in response to reading words with vertical semantics. This task is thus completely unrelated to the semantics of the words. We show that tongue movements are systematically modulated by verticality ratings collected from the same speakers. To investigate the source of the effect, we performed two additional, linguistically unrelated experiments. Experiment 2 demonstrates anti-phasic coupling between tongue body movements and vertical arm and leg movements. Experiment 3 investigates tongue body movements prior to head movements and uncovers preparatory tongue raising to head raising in contrast to head lowering. Taken together, the results indicate that the changes in *ja* associated with vertical semantics most likely emerge from anticipating a head movement in the direction of the spatial target associated with the read word in order to optimize the body position for subsequent actions. Thus, the results support the assumption that words reduce the uncertainty about future states of the world.

1   Introduction

The meanings that we glean from speech and writing have been shown to interfere with the execution of some physical responses. A movement in a direction that is incongruent with the spatial semantics of linguistic stimuli typically requires more time than movements that are directionally congruent [1–4]. For instance, such interference (facilitation of semantically congruent / inhibition of incongruent behaviour) has been observed for: button-press tasks [5]; tasks requiring eye-movements in response to lexical stimuli [6]; and tasks addressing the effect of rotational stimuli on sentence comprehension [2]. Other experiments have shown that words carrying spatial

information can evoke spontaneous congruent behaviour such as upward and downward saccades in response to words referring to the vertical dimension [7,8].

Van Elk et al. [9] pointed out that there are two possible explanations of how to interpret the results. While the results are often taken to providing evidence for a mental model of the world in which the semantics of the words are associated with typical actions such that a word's meaning is understood by using the internal model to mentally simulate a word's meaning. This view rests on a conceptualization in which meaning is somehow iconic. A more plausible explanation proposed by Van Elk et al. [9] is enactivism. According to this view, words simply serve to reduce hearers' uncertainty about future states of the world, and thus facilitate actions that prepare for them. This implies that the effects produced by a given word will exhibit a high degree of context sensitivity, rather than a uniform effect across task domains. Since previous experiments have focused on actions that are known to be directly involved in the exploration of space (head, arm, and eye movements), they provide evidence for both hypotheses. However, if semantics lead to anticipatory planning of moving specific body parts, and if this planning interacts with both context and the rest of the body, then an interesting possibility arises. Consistent modulatory differences ought to be observable in even highly automated muscle movements, such as those used in speech articulation. Importantly, these motor plans are not thought to be involved in more general spatial planning and hence are not thought to contribute to word meanings. To begin to address this question, we examine the basic downwards directed articulatory gestures that are exhibited by German speakers as they pronounce a high-frequency German word, *ja* ('yes'). Consistent with the foregoing, we predict that expectations evoked by linguistic stimuli will influence articulation.

Experiment 1 shares key properties with experimental designs typically used to demonstrate semantic priming effects on movement execution. However, it also contains changes that allow us to control for confounds such as the properties of stimulus material, task structure and subjects' attention to the task. In this modified verbal lexical decision task subjects were asked to respond to German words by uttering *ja* and *nein* ('no') in a manner completely unrelated to the semantics of the words, or to pronounce pseudowords. Crucially, the German words contained semantic information about locations and movements in the vertical dimension which were individually assessed in a separate experiment collecting verticality ratings from the same subjects.

2 Experiment 1

2.1 Methods

2.1.1 Subjects

Nine male and thirteen female native speakers of German (mean age 23.86; sd 2.9) participated in the experiment. None of the participants were wearing metal objects on or near their head that might have interfered with the electromagnetic recording of their articulatory movements.

2.1.2 Word Materials

The selected German words either carried information about location on the vertical dimension (e.g. under), denoted objects that occur in a typical location above or below the human eye level (e.g. airplane), or expressed a movement in the vertical direction (e.g. to climb). A set of 56 words associated with *up* (e.g. rocket) and a set of 56 associated with *down* (e.g. to fall) were balanced with respect to word class, length and frequency of occurrence in the SdeWaC corpus [10]. A third set of 56 pseudowords was

generated with the LINGUA pseudoword generator [11] trained on the 190,000 most frequent words in the SdeWaC corpus. These pseudowords matched the words with respect to capitalisation and word length.

For each word its frequency of occurrence in the SdeWaC corpus [10] was recorded and log-transformed for inclusion as a covariate in the statistical analysis. For each word, semantic similarity scores were calculated to each of the four words *ja* ('yes'), *nein* ('no'), *oben* ('up') and *unten* ('down'). Semantic similarity was estimated using the HiDEx vector-space model [12, 13]. In addition, subjects were asked to provide verticality ratings on a scale from -1 (down) to $+1$ (up) for both the words and the pseudowords (see below for further details). These (subject-specific) ratings were entered as a covariate in the statistical analysis.

2.1.3   Recordings

Articulatory movements of tongue and jaw were recorded at a sampling rate of 100 Hz with 16-channel device (WAVE, Northern Digital Inc.). Head movements were corrected for with the help of sensors on the nose bridge and left and right mastoids. A reference coordinate system was established by means of a 10 sec recording of three equilaterally placed sensors on a bite plate [14]. For the recording of articulatory movements, three sensors were attached to the tongue using PeriAcryl 90 HV dental glue: one approximately 3 mm behind the tongue tip (tongue tip sensor), one at the furthest point back on the tongue subjects could reach with their teeth (tongue body sensor) and one half way between these two (tongue mid sensor). The participant did not wear any metal objects on or near their head that might have interfered with the electromagnetic recording of their articulatory movements.

The audio signal was recorded with an Oktava MK-012 microphone (Sampling rate: 22.05 kHz, 32 bit) placed at a distance of 40 cm away from the subject. It should be noted that tongue position is corrected for head movement, i.e. analysed tongue positions quantify the relative position to the skull and not the absolute distance from the floor. Furthermore, the apparatus only allowed for recordings of the joint movement of tongue and jaw. As it is the joint effect of both movements that determine vowel quality and the subject performed her head movements with her mouth closed, no further attempts were made to tease apart jaw movements from tongue movements.

2.1.4   Procedure

Subjects performed a modified verbal lexical decision task with lateral presentation position priming. A trial was initiated with a fixation cross presented for 750 ms in the centre of the screen, followed by an arrow pointing left or right for 1250 ms also placed in the centre of the screen, followed in turn by the stimulus presented for 2500 ms. Fixations cross and arrows had a width of 3 cm. The stimulus was placed either in the centre of the left half of the screen or in the centre of the right half of the screen (18.7"x11.6" LCD screen,120 Hz, resolution of 1680x1050 pixels). All stimuli were shown in black on a light grey background using the monospaced font Courier New ensuring letter height of 2.2 cm for capitalised letters. In case a real word appeared on the side indicated by the arrow, subjects had to utter *ja*, otherwise they were requested to say *nein*. In case a pseudoword appeared, subjects were asked to read this pseudoword out loud.

The word stimuli were divided into pairs of two up words or two down words such that each pair was matched as well as possible for word class, length, and frequency of occurrence (e.g. *tief* and *niedrig*, deep and low, *Gipfel* and *Gebirge*, summit and mountains). For each pair, one of the words was assigned to the *ja* response (left arrow and left presentation (LL), or right arrow and right presentation (RR)) and one to the

nein response (left arrow and right presentation (LR), or right arrow and left   106
presentation (RL)). The sets of words and pseudowords were merged and placed in   107
random order. A copy of this list was created in which the yes-no-assignment within   108
pairs was reversed. From the resulting two lists, we constructed versions with reversed   109
item order, resulting in a total of four lists. For a given subject, the words were   110
randomly assigned to either the LL or the RR condition for *ja* responses, and the LR or   111
the RL condition for *nein* responses. This procedure resulted in an equal number of *ja*   112
and *nein* responses to each German word while ensuring counterbalancing.   113

Subjects performed a modified verbal lexical decision task where the task was   114
designed to direct their attention away from the vertical dimension and on to the   115
horizontal dimension. Following a fixation cross, participants saw an arrow pointing   116
either left or right. This arrow was followed in turn by the stimulus, which was placed   117
either in the left half of the screen or in the right half. When an actual German word   118
was shown on the side indicated by the arrow, participants were requested to say *ja*, and   119
if the word appeared on the other side of the screen, they were asked to respond with   120
nein ('no'). When a pseudoword appeared, it had to be read out loud. By having   121
subjects read out loud non-existent words, we sought to avoid motor priming effects   122
from previous *ja* responses as well as conceptual priming effects from the previously   123
presented word. Accordingly, we predicted modulations of the *ja* responses specifically   124
when following a trial on which a pseudoword was pronounced. In the obvious interest   125
of producing a natural task design, we both elicited *ja* and *nein* responses. However,   126
nein responses were never intended to be an independent variable because the   127
articulation of *nein* has a more complex, and yet shallower trajectory than *ja*, and   128
whereas *ja* requires a simple lowering movement, *nein* combines downward, upward, and   129
fronting movements of the articulators.   130

Importantly, the complexity of the experiment guaranteed that subjects were fully   131
engaged in the task such that the influence of high-level cognitive control mechanisms   132
on the automated responses were reduced to a minimum. Indeed, only one of the   133
subjects reported that she had noticed words to carry up / down semantics when asked   134
after the experiment.   135

A separate experiment collected ratings for verticality from the same subjects after   136
the tongue movements were recorded.   137

The experiment was initiated with a training session with 20 items following which   138
participants were given the opportunity to ask questions of clarification. The test words   139
were presented in two blocks of 84 test items (56 words, 28 pseudowords) each of which   140
was preceded by five practice items. A short break separated the two blocks. The total   141
duration of the experiment was approximately 20 min. After completion of this   142
experiment, subjects were asked to provide a verticality rating first for the pseudowords,   143
and subsequently for the actual German words, using a continuous sliding scale from -1 144
(completely down) to $+1$ (completely up).   145

2.1.5   Preprosessing   146

The timestamps for the *ja* responses and *nein* responses were extracted from the audio   147
file using an automatic aligner [15] and manually corrected were necessary using   148
Praat [16] (Version 5.3.41). These timestamps were then used to extract the time series   149
of sensor positions from the articulography record. Data from incorrect responses, as   150
well as data from trials where sensors got disconnected and had to be glued back onto   151
the tongue, were removed prior to analysis (6.6% data loss equally distributed across   152
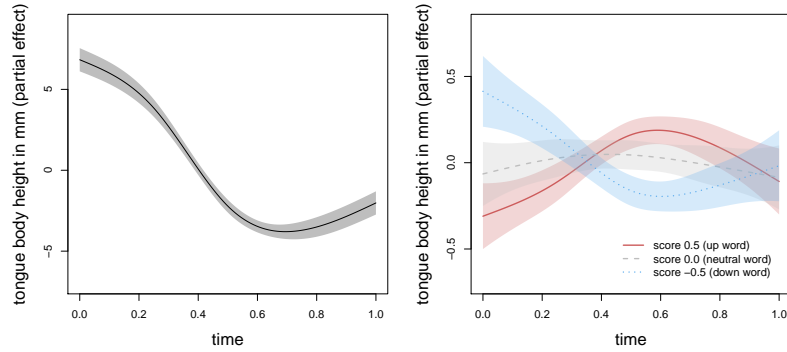down (51%) and up (49%) words).   153

Fig 1. Vertical position of the tongue body sensor. Left panel: partial effect for tongue body height as a function of normalised time; Right panel: partial effects for 'up' words (verticality rating score 0.5), neutral words (verticality rating score 0), and 'down' words (verticality rating score -0.5). The partial effects in the right panel modulate the main effect in the left panel, rendering the articulatory trajectory more steep for the 'down' words and more shallow for the 'up' words.

2.1.6 Statistical Analysis

The statistical analysis was conducted in R (3.3.1) using Generalized Additive Mixed Models (GAMM) (mgcv package (1.8-14), [17–20]). Our hypothesis was that after having read out a pseudoword at the preceding trial, subjects would show differences in their articulation of *ja* depending on whether they read words with a negative or positive verticality rating. We tested this hypothesis by examining a three-way interaction of time t (normalised between 0 and 1), rating score r, and a factor f specifying whether the preceding trial required the reading out loud of a pseudoword. This three-way interaction was modelled by means of tensor product regression splines for the two levels of f. Time varying random effects for participant and item were brought into the model by means of factor smooths [21]. Changes in articulation that unfolded in the course of the experiment were modelled by means of a thin plate regression spline for standardised trial number z and by means of a three-way interaction of t, z and f also modelled with tensor product splines. Temporal autocorrelation in the residual errors was captured by means of an AR(1) process with proportionality parameter $\rho = 0.95$. No effect was observed for frequency of occurrence or vector-space similarity measures.

Since the articulation of *ja* is primarily carried out by the tongue blade [22] whose movements we captured with the tongue body sensor we focus our analysis on this sensor. Nevertheless, since the entire tongue is involved in the articulation, we also analysed the tongue mid and tongue tip sensors for completeness.

2.2 Results

As anticipated, articulation was affected in those trials for which there was no articulatory or conceptual priming from the preceding trial. Following a pseudoword trial the position of the tongue body sensor varied with vertically rating ($F_{(9.237,11.872)} = 2.623, p = 0.00194$) in the expected direction (Figure 1). Remarkably, when speakers say *ja* in response to words with strong down semantics (e.g. *floor, deep*) they respond with steeper articulatory trajectories that started higher and moved further down as compared to words with strong up semantics (e.g. *sky, sun*). When words have strong up semantics, the articulatory trajectory becomes more shallow. Crucially, the articulatory pattern was unaffected by neutrally rated words. Similar and

significant effects were observed for the tongue tip and tongue mid sensors (see Supplementary Materials).

When the preceding item was not a pseudoword, there was no effect of verticality on the tongue movement ($F_{(6.629, 8.703)} = 1.459, p = 0.14807$). Analysis of error rates ($\chi^2_{(1)} = 0.04$) and response duration ($t_{(1128)} = -0.36$) did not reveal any significant differences between the two conditions. Further analyses ruled out semantic, frequency and co-articulation effects induced by the presented items (see Supplementary Materials).

2.3   Discussion of experiment 1

This experiment sought to test the hypothesis that even highly automated movements could be influenced by semantic information even when these movements were not directly related to the semantics carried by the words. In accordance with this hypothesis, we found that simple downward movements of the tongues of participants saying *ja* could be modulated by words carrying information about vertical locations or displacements.

Previous experiments addressing semantic interference on movement have shown that linguistic inputs associated with spatial semantics that are incongruent with the direction of evoked movements produce slower responses and higher error rates [1,6]. By contrast, here the response durations and error rates do not differ between the incongruent and congruent conditions. This is exactly as expected for an effect that is not driven by explicitly requesting participants to execute motions in a pre-specified direction. We did not ask people to raise or lower their tongues to indicate their responses. We simply asked them to respond with *ja* or *nein* during a task that drew their attention to the horizontal dimension.

Moreover, although word semantics here are potentially confounded in their vertical and emotional dimension (i.e. down is sad, up is happy), studies using electromagnetic articulography have shown that the differences in velocity and displacement of the tongue, jaw and low lip are far from aligned with emotional valences. Specifically, shallower movements are observed when dealing with sad as compared to positive emotions [23,24] whereas here we observed the opposite: our participants articulated *ja* with shallower movements after the up words.

Further, the literature has often taken findings showing interference between movements and semantics as resulting from mental simulation (involving the retrieval of prior experiences, e.g. looking up to the sun, or looking down to the ground). However, while these experiments deal with task-induced movements that are likely to be more frequently expressed in the presence of either up or down semantics [25], *ja* is not uttered more frequently in response to down (or sad) words and less frequently in response to up (or happy) words: google document frequencies for English point to exactly the opposite pattern (1.370 million hits for *yes* and *up*, 379 million hits for *yes* and *happy*, 619 million hits for *yes* and *down* and 118 million hits for *yes* and *sad*).

This poses the question which processes actually give rise to the modulation of the articulatory trajectory observed. The only dimension in which the semantics of our stimulus words consistently differed was that of vertical directionality. One possible explanation of the observed attenuation and amplification of *ja* trajectories is that of a representation of directionality on which the brain draws when making sense of input from the world and its own body, and which mediates between input and output resulting in conflict or facilitation depending on whether both align in this dimension or not. The enhanced downward trajectory following down words would then arise as a consequence of resonance between the initial motor simulation elicited by the content words shown and subsequent articulatory planning for *ja*.
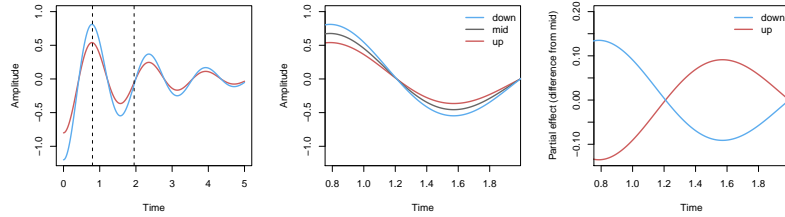
Fig 2. Left panel: Two damped harmonic oscillations with equal dampening but different intercepts at t = 0 (blue: low onset, red: high onset). Mid panel: Magnification of the oscillations between the dashed lines in the left panel. The average between the red (up words) and the blue line (down words) is represented in black. Left panel: Differences of the two oscillations from the mean.

Furthermore, if word meanings are ultimately grounded at least in part in motor simulations, then this would predict that the effects on tongue height observed in experiment 1 should replicate in other tasks where the relevant factors are held constant. Specifically, if the up words elicit consistent simulations for upwards movements across different body parts one would expect the same effect on tongue position irrespective of which limb is involved.

To test whether this is in fact the case, we conducted a further two experiments that involved vertical leg movement (experiment 2a) and arm movements (experiment 2b) and their consequences on tongue position.

However, there is an alternative way of understanding the present results that take as point of departure the similarities of the physics of tongue movements to those of a pendula driven by torsion springs [26]. When a spring is extended or pushed down, a dampened harmonic oscillation is created as shown in the left panel of Fig 2. The amplitude of the oscillation depends on the extent to which the spring is stretched or compressed, represented by the offsets or intercepts at t = 0. When this offset is smaller, so is the force applied to the spring is smaller and as a result, the amplitude of the oscillation will be reduced, as illustrated by the red oscillation[1]. The oscillations between the two vertical dashed lines are magnified in the second panel of Fig 2. The average of the two oscillations is shown in black. When we plot the difference of the two oscillations from this average, the curves shown in the right panel are obtained. These curves represent the partial effect of offset on the shape of the curve. These partial effects mirror those observed in experiment 1 (Fig 1). If the understanding of an up or down word resulted in an initial vertical offset of the tongue, with a higher intercept for up words at $t = 0$ before the initiation of the articulation of 'ja', then the observed pattern in Fig 1 follows straightforwardly.

We hypothesize that the different offsets for up and down words arise as the consequence of the different head movements that would be required to verify the presence of the referents denoted by these words. It is well established that motor systems can be activated even though they are not required for intentional action. For instance, just listening to speech leads to muscle tension in the articulators and even to muscle tension in the upper body [29]. As shown by Shaw (1938) [30], the imagining of

---

[1] Following Simko and Cummins (2009) [26], we used the solution $x = e^{-\gamma t} * (x_0 cos(\omega t) + x_0 \gamma \frac{sin(\omega t)}{\omega})$ to the linear differential equation for a damped harmonic oscillator, $\ddot{x}(t) + 2\gamma \dot{x}(t) + \omega_0^2 x(t) = 0$, where ω represents the frequency of the oscillation, γ represents the strength of dampening, and x_0 represents the amplitude at $t = 0$. We are fully aware that speech is produced by a complex, bio-mechanically and task-dynamically coupled system [27, 28]. The movement pattern of single gestures, however, especially one that describes the raising of the tongue body during the production of 'ja', can be approximated by this formula.

actions gives rise to the tensing of various muscle systems, not just those that would be immediately involved in these actions. If the participants in Experiment 1 engaged to some extent in imaging of the referents of the up and down words they were presented with, this may have resulted in changes in the muscle systems typically active during perception of these referents, i.e., those involved for upward head movements in the case of up words and those involved for downward head movements in the case of down words. Experiment 3 tests this hypothesis by examining how the word 'ja' is articulated before the execution of an upward vs. downward head movement.

3   Experiment 2a: Leg movements

3.1   Methods

3.1.1   Procedure

To test how tongue movements in the vertical dimension interact with the physical state of other body parts, one 23-year-old female native speaker of German participated in an experiment investigating the effect of leg movement on the resting state of the tongue. On each of 25 pseudo-randomized trials an arrow appeared on the screen pointing either left or right. At the same time a low beep (inaudible to the participant) was registered on the audio recording device. The subject was instructed to lift and stretch out her left or right leg respectively, and to lower it upon the presentation of a green circle 2 s later. A high beep (inaudible to the participant) was simultaneously sent to the recording device. The recording procedure for articulatory movements used in experiment 1 was also implemented for this experiment. The recordings of the articulatory time series and of the beeps were synchronized. The latter were used to segment the articulatory time series into periods with down and up leg movements.

3.1.2   Statistical Analysis and Results

The upper panel of Fig 3 summarizes the change in the vertical position of the tongue body sensor over normalized time as legs were raised and lowered. The red dashed vertical line specifies the point in normalized time at which the instruction was given to lower the leg. The median vertical position of the tongue body sensor over time was estimated with a thin plate regression spline smooth fitted with a quantile GAM (henceforth QGAM), using the implementation of the R package **qgam** ($F_{(2.987,3.694)} = 23.27, p < 0.001$). QGAMs [31], which integrate quantile regression [32] with the generalized additive model (GAM) [17–20, 33], are distribution-free. Hence, in contrast to Gaussian GAMs, autocorrelations in the residuals do not require remedial action (as implemented for experiment 1 with an AR(1) process in the errors; see Baayen et al. [34] for discussion of pros and cons of modeling with AR(1) noise). The model indicates that the tongue body enters into a movement that is antiphasic with that of the leg. Experiment 2b investigates whether a simple antiphasic is present for arm movements as well.

4   Experiment 2b: Arm movements

4.1   Methods

4.1.1   Procedure

The same subject performed a parallel experiment in which either the left or the right arm was moving a 1 kg or 2 kg weight up and down. The subject was requested to keep
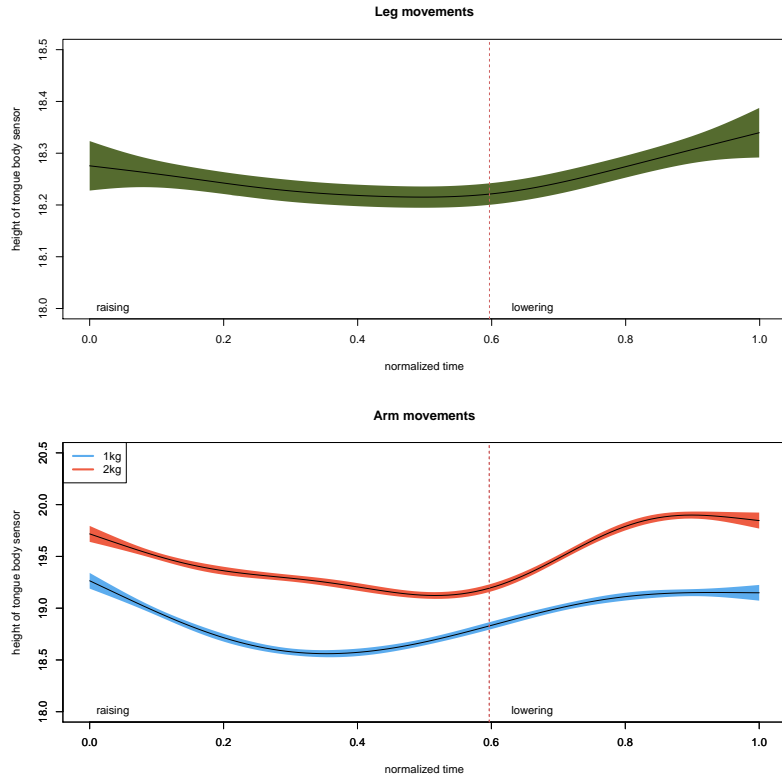
Fig 3. Tongue body height as a function of normalized time. Upper panel: Vertical position of the tongue body sensor during leg movement (experiment 2a). Bottom panel: Vertical position of the tongue body sensor during arm movement (experiment 2b). The dashed vertical red line indicates the moment in time when the instruction for lowering leg or arm was given. The tongue body sensor moved down while limbs were raised, and moved up while limbs were lowered.

the arm stretched and the plane of movement was parallel to the midsagittal plane. In the first 25 trials, the subject lifted the 1 kg weight, and in the remaining 25 trials she lifted the 2 kg weight. Otherwise the procedure was exactly the same as for experiment 2a.

4.1.2 Statistical Analysis and Results

A QGAM was fitted to the articulatory trajectory of the tongue body sensor, with weight as additional predictor. This predicter interacted with time, as illustrated by the smooths in the lower panel of Fig 3.

The overall position of the tongue was higher when lifting 2 kg weights compared to lifting 1 kg weights ($\beta = 0.59, sd = 0.013, z = 46.63, p < 0.001$). For the 2 kg the minimum of the articulatory trajectory ($F_{(4.903, 4.995)} = 888.1, p < 0.001$) is close to the point in time at which the instruction for lowering the arm was received. For the 1 kg weight the minimum of this trajectory ($F_{(4.750, 4.969)} = 634.0, p < 0.001$) is located somewhat earlier. As in experiment 2a, the general pattern is one in which the tongue body sensor oscillates in an antiphasic manner with arm movement in the vertical dimension.

4.2 Discussion

Experiments 2a and 2b underline the degree to which the various muscle systems in the body are systematically interconnected, [35], albeit with stronger interconnections for the arms compared to the legs. The tongue itself is embodied. Importantly, the antiphasic oscillation of tongue body and limbs are not straightforwardly reconcilable with an iconic theory of embodied meaning. If embodied meaning for upwards movement or high location (lift, climb, reach) were to lead to simulated motor responses for up movement in the limbs, experiments 2a and 2b lead to expect a downward movement of the tongue body, rather than an upward tongue movement expected under iconicity.

Importantly, although embodied iconicity can explain the results of experiment 1, as preceeding up words lead to muted downwards tongue body movement during the articulation of 'ja', experiments 2a and 2b show that motor simulation of upwards movements of arms and legs actually should lead to more pronounced downwards movements of the tongue body following up words. Since, as Van Elk et al. [9] has pointed out, it is unclear what the relevant motor simulations would be for words like *ceiling* and *floor*, a standard interpretation of iconicity between motor movement and vertical semantics can be ruled out. In the next section, we consider our alternative explanation which predicts anticipatory tongue movements depending on whether one has to look up or down to verify whether the referents denoted by an up or down word is present.

5   Experiment 3

Up and down movements of the head are likely to have consequences for the position of the tongue body. The tongue muscle that is responsible for backwards and upwards movement of the tongue, the styloglossus, is attached to the styloid processes, two long and thin pointed pieces of bone that protrude from the skull just below the ear. When the head is raised, the styloid processes move back, pulling on the styloglossus with as a result that the tongue body is raised. Thus, for purely physiological reasons, we expect that raising the head will be reflected in an upwards movement of the tongue. Experiment 3 tested this prediction by requesting the same subject participating in experiments 2a and 2b to raise and lower her head in response to a verbal cue. In order to clarify the hypothesized effect of anticipated head movements on articulation, we instructed the subject to say the word 'ja' upon presentation of the verbal cue and to move her head upwards or downwards immediately after 'ja' offset.

5.1   Method and procedure

On each of 25 pseudo-randomized trials an arrow pointing up or down was presented in the middle of the screen. The subject waited for two seconds (preparation phase) until the question *Bereit?* (engl. ready?) appeared on the screen. The subject was instructed to reply with *ja* ('yes') (ja phase), and subsequently move her head as far as possible in the indicated direction (head movement phase). The experiment was carried out with the same setup as experiment 2. As before, acoustic triggers (inaudible to the participant) were sent to the recording device to mark the temporal boundaries between the three phases.

5.2   Results and Discussion

A QGAM predicting tongue body position from time and instructed direction of head movement revealed the expected upward swing of the tongue body when the head was moved up. Unsurprisingly, given the narrow confidence intervals, both curve are well
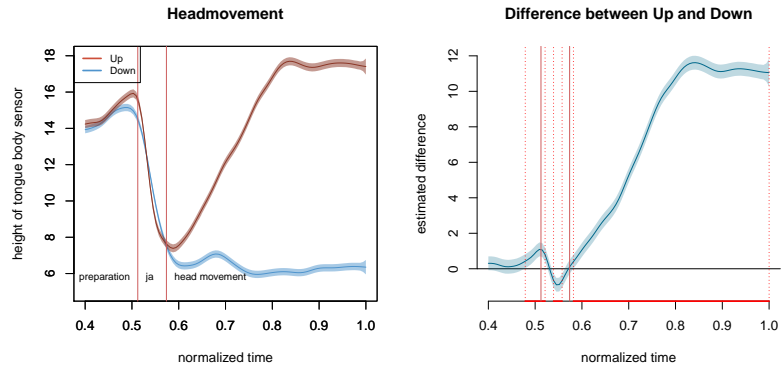
Fig 4. Tongue body height as a function of normalized time in experiment 3. Left panel: estimated vertical tongue trajectory in down (blue) and up (red) trials. Right panel: Difference between down and up trials. Intervals of time during which the two articulatory trajectories were significantly different (at $\alpha = 0.001$) are highlighted in red on the x-axis.

supported statistically (up: $F_{(33.08,40.28)} = 44133, p < 0.0001$; down: $F_{(42.00,49.50)} = 23680, p < 0.0001$). The difference between the articulatory trajectories for up and down movement is shown in the right panel. This panel highlights the time intervals during which the two curves are significantly different (at $\alpha = 0.001$).

During the preparation phase (normalized time 0 to 0.4) the vertical position of the tongue body sensor was basically the same irrespective of the head movement the subject was requested to carry out. After the initiation of the head movement, the vertical position of the tongue body sensor was consistently and substantially higher for the upward movements compared to downward movements, as expected for the physiological reasons outlined above. Of special interest is the difference in the tongue body sensor position before and during the articulation of 'ja'. When instructed to raise the head, the tongue body sensor moves upwards slightly before the onset of articulation, compared to when a downward head movement was requested. During the articulation of 'ja', tongue height in the upwards condition is initially higher than that in the downwards condition. Subsequently, it moves below the vertical position characterising the downwards condition, to finally end at approximately the same low position at which the head movement is initiated.

The two key findings of interest are, first, that an upward head movement indeed induces tongue body raising, and second, crucially, that tongue body raising is present before the onset of articulation, and long before the actual head movement is executed. These results lead to the following interpretation of experiment 1. After having understood the up and down words, an anticipatory raising of the tongue body is taking place as a consequence of the imaging of the body posture that is typical for when the referents of these words are perceived [29, 30]. In experiment 3, we induced this anticipatory muscle tension not by means of prior presentation of up or down words, as in experiment 1, but by directly requesting the participant to move her head up or down, albeit after a delay during which the word 'ja' had to be articulated. In the hypothesized dampened harmonic oscillation depicted in the left panel of Fig 2, this anticipatory raising is represented by the higher offset for the up words at $t = 0$. For the articulation of 'ja', the tongue has to be raised towards the alveolar ridge which is the starting point for the second and third panels of Fig 2. For the articulation of the vowel, the tongue moves down and due to the spring like properties of the tongue, the downwards trajectory reaches a lower minimum for the down words compared to the up

words. 403

6 General Discussion 404

Experiment 1 documented that subsequent to reading a word with up semantics, the 405
articulation of 'ja' was executed with shallower downward movement trajectories 406
compared to when a word with down semantics was read. If this effect arises due to 407
iconicity between verticality semantics and tongue movement, such that down words 408
induce stronger downward tongue movement, then one would expect that the tongue 409
would move upwards with upward limb movements and downward with downward limb 410
movements. Experiments 2a and 2b tested this prediction and showed the opposite. 411
When arms or legs are moved upwards, the tongue body moves downwards. When arms 412
or legs are moved downwards, the tongue body moves upwards. Incidentally, 413
experiments 2a and 2b demonstrate that the tongue is not decoupled from the rest of 414
the body, but is part of a systematically interconnected muscle system [35]. 415

Experiment 3 moved from limb movements to head movements. When the head 416
moves upwards, the tongue is pulled back and up by the styloglossus muscle. In this 417
case, the physical reasons for why the tongue moves in the way it does are much clearer 418
than for experiments 2a and 2b. Crucially, experiment 3 also documents that when the 419
speaker was articulating 'ja' and anticipating an upward head movement, the tongue 420
body moved slightly up compared to when the speaker was anticipating a downward 421
head movement. This anticipatory tongue movement preceeded in time both the onset 422
of articulation and the actual head movement that was initiated after 'ja' had been 423
pronounced. 424

Given the results of these three experiments, it is clear that the stronger downward 425
movement of the tongue following the presentation of down words is unlikely to be due 426
to some form of mental similation of the words' referents or to some kind of motor 427
iconicity, according to which a downward movement of the tongue would be in sync 428
with simulated downward movements of body parts (see for example [1–4] for 429
experimental effects of directional congruency). Apart from the problem that is difficult 430
to see what motor simulations would consist of for words as diverse as 'star', 'airplane' 431
and 'basement', experiments 2a and 2b clearly demonstrate that downward and upward 432
limb movements affect the tongue body position anti-iconically instead of iconically. 433
This leaves us with two alternative explanations. 434

First, it is conceivable that the affordances of the up and down words' referents are 435
at issue. In order to make use of an airplane one normally has to look up and to pick up 436
something from the floor one usually has to look down. This leads to the hypothesis 437
that it is the head movements typically prompted by the up and down words' referents 438
that are anticipated by the subjects in experiment 1, just as the head movements in 439
experiment 3 are anticipated long before the actual head movement is actually carried 440
out. In other words, it is the words' meaning that is informative about changes in the 441
environment and future body states and so causes anticipatory motor plans [9]. Up and 442
down words put the predictive system into differentiable states reflecting the 443
preparedness for different sets of actions. In this light, the voluntary saccades observed 444
by Demerais and Cohen [7] and Spivey and Geng [8] reflect the preparedness of the 445
hearer to visually explore the environment directed by the linguistic input to allow for 446
subsequent interaction with it. Further, Pelekanos and Moutoussis [36] reported 447
language to influence visual contrast sensitivity, and Mathot et al. [37] showed that 448
words referring to brightness cause changes in pupillary responses. In the light of this 449
explanation, the main difference between experiment 1 and experiment 3 is that in 450
experiment 3 a head movement is first anticipated and eventually carried out, whereas 451
in experiment 1 it is anticipated but subsequently inhibited [38]. 452

Alternatively, instead of anticipating future adequate action, it is possible that understanding an up or down word builds on memories of the state of the body that is typical for encounters with the referents of these words. It is well known that the imaging of body movements gives rise to tensing of various muscle systems [30, 38–40]. Possibly, up and down words activate memories of the muscle configurations for the head movements involved when perceiving and interacting with their referents (see [41] for the importance of mental practice for motor skills).

What the two explanations have in common is that a simulation of the meaning of the up or down words, what ever that might be, is not at issue, but rather the state of the body that is typical during or following experiences with the referents of these words. It is in this light that the modulation of tongue movements by lexical semantics, as observed in experiment 1, is best understood. Even when highly-practised motor commands are being executed, as is the case for German *ja*, the most frequent word in the KEC corpus of German conversational speech [42], we see evidence of motor imaging or anticipatory motor planning.

Supplementary Materials

Datasets, materials, and analyses can be found online `https://osf.io/7wdpx/`.

Acknowledgements

Contributions

Conceptualization: FT MR HB BK; Data curation: FT MR HB; Formal analysis: FT HB; Funding acquisition: HB; Investigation: FT MR ST; Methodology: FT MR ST; Project administration: MR HB; Resources: HB; Software: FT; Supervision: MR HB; Validation: FT HB; Visualization: FT HB; Writing – original draft: FT MR HB; Writing – review and editing: FT MR HB BK

Declaration of Interest

The authors declare that they have no competing financial interests.

References

1. Glenberg AM, Kaschak MP. Grounding language in action. Psychonomic bulletin & review. 2002;9(3):558–565.

2. Zwaan RA, Taylor LJ. Seeing, acting, understanding: motor resonance in language comprehension. Journal of Experimental Psychology: General. 2006;135(1):1.

3. Barsalou LW. Grounded cognition. Annu Rev Psychol. 2008;59:617–645.

4. Fischer MH, Zwaan RA. Embodied language: A review of the role of the motor system in language comprehension. The Quarterly Journal of Experimental Psychology. 2008;61(6):825–850.

5. Lachmair M, Dudschig C, De Filippis M, de la Vega I, Kaup B. Root versus roof: automatic activation of location information during word processing. Psychonomic bulletin & review. 2011;18(6):1180–1188.

6. Dudschig C, Souman J, Lachmair M, de la Vega I, Kaup B. Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. PloS one. 2013;8(2):e56872.

7. Demarais AM, Cohen BH. Evidence for image-scanning eye movements during transitive inference. Biological Psychology. 1998;49(3):229 – 247.

8. Spivey MJ, Geng JJ. Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. Psychological Research. 2001;65(4):235–241.

9. Van Elk M, Slors M, Bekkering H. Embodied language comprehension requires an enactivist paradigm of cognition. Frontiers in Psychology. 2010;1:234.

10. Faaß G, Eckart K. SdeWaC–a corpus of parsable sentences from the web. In: Language processing and knowledge in the Web. Springer; 2013. p. 61–68.

11. Westbury C, Hollis G, Shaoul C. LINGUA: the language-independent neighbourhood generator of the University of Alberta. The Mental Lexicon. 2007;2(2):271–284.

12. Shaoul C, Westbury C. Exploring lexical co-occurrence space using HiDEx. Behavior Research Methods. 2010;42(2):393–413.

13. Lund K, Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods Instruments and Computers. 1996;28(2):203–208.

14. Tomaschek F, Wieling M, Arnold D, Baayen RH. Frequency effects on the articulation of German i and u: evidence from articulography. In: Proceedings of Interspeech, Lyon; 2013. p. 1302–1306.

15. Rapp S. Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models. An aligner for German; 1995.

16. Boersma P, Weenink DJM. PRAAT, a system for doing phonetics by computer, version 3.4. Institute of Phonetic Sciences of the University of Amsterdam, Report. 1996;132.

17. Wood SN. Generalized Additive Models. New York: Chapman & Hall/CRC; 2006.

18. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B). 2011;73:3–36.

19. Wood SN. On p-values for smooth components of an extended generalized additive model. Biometrika. 2013;100:221–228.

20. Wood SN. A simple test for random effects in regression models. Biometrika. 2013;100:1005–1010.

21. Baayen RH, Vasishth S, Bates D, Kliegl R. The cave of Shadows. Addressing the human factor with generalized additive mixed models. Journal of Memory and Language. to appear;.

22. Fowler C. Coarticulation and theories of extrinsic timing. Journal of Phonetics. 1980;8:113–133.

23. Lee S, Yildirim S, Kazemzadeh A, Narayanan S. An articulatory study of emotional speech production. In: Interspeech; 2005. p. 497–500.

24. Lee S, Narayanan S. Effects of Emotion on the Lower Lip Movements at Phrase Boundaries. Proceedings Of Speech Prosody, Shanghai, Cn. 2012;.

25. Hutchinson S, Louwerse MM. Language statistics explain the spatial–numerical association of response codes. Psychonomic bulletin & review. 2014;21(2):470–478.

26. Simko J, Cummins F. Sequencing embodied gestures in speech. Advances in Cognitive Systems. 2009;.

27. Saltzman EL, Munhall KG. A Dynamical Approach to Gestural Patterning in Speech Production. Ecological Psychology. 1989;1(4):333–382.

28. Fowler C, Saltzman E. Coordination and Coarticulation in Speech Production. Language and Speech. 1993;36(2-3):171–195.

29. Smith AA, Malmo RB, Shagass C. An electromyographic study of listening and talking. Canadian Journal of Psychology/Revue canadienne de psychologie. 1954;8(4):219.

30. Shaw WA. The distribution of muscular action potentials during imaging. The Psychological Record. 1938;2(6):195–216.

31. Fasiolo M, Goude Y, Nedellec R, Wood SN. Fast calibrated additive quantile regression; 2017. Available from: https://github.com/mfasiolo/qgam.

32. Koenker R. Quantile regression. 38. Cambridge university press; 2005.

33. Hastie TJ, Tibshirani RJ. Generalized Additive Models. London: Chapman & Hall; 1990.

34. Baayen RH, van Rij J, de Cat C, Wood S. Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In: Mixed-Effects Regression Models in Linguistics. Springer; 2018. p. 49–69.

35. Myers TW. Structural integration—developments in Ida Rolf's 'recipe'—Part 3. An alternative form. Journal of Bodywork and Movement Therapies. 2004;8(4):249–264.

36. Pelekanos V, Moutoussis K. The effect of language on visual contrast sensitivity. Perception. 2011;40(12):1402–1412.

37. Mathot S, Grainger J, Strijkers K. Pupillary Responses to Words That Convey a Sense of Brightness or Darkness. Psychological Science. 2017;28(8):1116–1124. doi:10.1177/0956797617702699.

38. Guillot A, Di Rienzo F, MacIntyre T, Moran A, Collet C. Imagining is Not Doing but Involves Specific Motor Commands: A Review of Experimental Data Related to Motor Inhibition. Frontiers in Human Neuroscience. 2012;6:247.

39. Jeannerod M. The representing brain: Neural correlates of motor intention and imagery. Behavioral and Brain Sciences. 1994;17(2):187–202.

40. Bonnet M, Decety J, Jeannerod M, Requin J. Mental simulation of an action modulates the excitability of spinal reflex pathways in man. Cognitive Brain Research. 1997;5(3):221 – 228.

41. Feltz DL, Landers DM. The Effects of Mental Practice on Motor Skill Learning and Performance: A Meta-analysis. Journal of Sport Psychology. 1983;5(1):25–57.

42. Arnold D, Tomaschek F. The Karl Eberhards Corpus of spontaneously spoken southern German in dialogues — audio and articulatory recordings. In: Draxler C, Kleber F, editors. Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum. Muenchen: Ludwig-Maximilians-Universitaet; 2016. p. 10–13.