Rethinking Best Match Using Movement

Andrew A. Hedding

1. Introduction

Syntactic Agreement is often assumed to be subject to constraints on locality, such as Relativized Minimality (Rizzi 1990) or Attract Closest (Chomsky 2000). In essence, these constraints state that if some probe is looking for a feature [F], then it will Agree with the most local goal that bears that feature (1). Skipping a local goal that bears that feature in favor of a non-local goal is not possible (2).

(1) Locality Satisfied

\[ \begin{array}{ccc}
X & Y & Z \\
\end{array} \]

(2) Locality Not Satisfied

\[ \begin{array}{ccc}
* & X & Y \\
\end{array} \]

While there are many phenomena that support the idea that Agreement is a fundamentally local relationship, there exist patterns of morphological agreement which appear to be sensitive to other factors beyond the structural relationship of potential goals. In many languages, morphological agreement indexes either the subject or the object, depending on where they fall on some person hierarchy. A pattern of this type can be seen in Algonquin, a dialect of Ojibwe. Morphology on the verb (underlined) will index a local person that is a subject (3a) or object (3b), when it co-occurs with a third person proximate argument.

(3) a. ni-wa:bam-a:-ina:n
    1-see-3obj-1pl
    ‘We see her.’

If, however, a clause contains two third person arguments—one proximate and one obviative—then agreement will mark the proximate argument, whether it is subject (4a) or object (4b).

(4) a. o-wa:bam-a:-wa:-an
    3-see-3obj-3pl-3
    ‘They see the other.’

Importantly, this pattern suggests that third person proximate arguments are potential targets for Agreement in the language, but that local person arguments are a “better” goal, in some sense.

Patterns of this kind have often been analyzed using probes that are relativized to multiple features (e.g. Béjar, 2003; Béjar & Rezac, 2009; Deal, 2015). For instance, Oxford (2019) argues that the underlined morphology in (3-4) indexes Agreement with a probe on Infl which is looking for a set of features [uperson-uproximate-uparticipant]. Given this relativization, a local person will completely satisfy the needs of this probe (5a), a proximate argument will mostly satisfy its needs (5b), and an obviative person will partially satisfy its needs (5c).

*Andrew A. Hedding, University of California, Santa Cruz, ahedding@ucsc.edu. I am very grateful to Roselia Durán Cruz, Erclacio Gracida Cruz, Juan Gracida Ortiz, Irma López Basurto, one additional consultant, and most especially Natalia Gracida Cruz for generously sharing their knowledge with me by providing the judgements presented here. I also thank Jack Duff, Ben Eischens, Jess H.K. Law, Ur Shlonsky, Ivy Sichel, Maziar Toosarvandani, and audiences at UC Santa Cruz, UC Berkeley, and WCCFL 39 for helpful feedback.
Intuitively, a probe “prefers” to agree with a local person when it can, given that a local person will completely satisfy its needs. Agreement with non-local persons—while not optimal—is possible when there is no local person in the derivation.

This approach, however appealing, introduces a tension into our understanding of syntactic Agreement: how do probes navigate the potentially conflicting requirements of agreeing with the goal that will best satisfy their needs, while still agreeing locally? Is locality respected even in cases when the most local goal only partially satisfies the probe (6a), or can a probe skip a local goal if there is some non-local goal that will check off more of its features (6b)?

Some previous work has attempted to alleviate this tension using a constraint called Best Match (Coon & Bale 2014, Oxford 2014, van Urk 2015, Oxford 2019). Though implementations differ, Best Match constraints share the intuition that a probe will agree with the goal that bears the most features that it is relativized to. However, one point of debate in previous work on Best Match is the way that it interfaces with locality constraints. According to Oxford (2019), Best Match only allows a probe to decide between two goals which are equidistant. He argues that objects undergo movement to spec-vP in Algonquin, and thus assumes that subjects and objects are equidistant to the probe which triggers the underlined agreement morphology in (3-4). Coon & Bale (2014) and van Urk (2015), on the other hand, argue that Best Match can trigger Agreement with a non-local goal in certain circumstances. Specifically, they propose that a probe can skip a local goal if there is a non-local goal that better matches its features.

In this paper I propose a reformulation of the Best Match constraint using data from a pattern of Ā-movement in San Martín Peras Mixtec (SMPM). Though both wh-words and foci must move to the same left-peripheral position in the language, the movement probe on C will agree with and attract a wh-word across a more local focus. Throughout this paper, I refer to this pattern as the Wh-over-Focus Preference.
Furthermore, I assume that if a probe triggers movement, a copy of the goal that values the probe will be internally merged into its specifier (Chomsky 2000). I propose that matching with a probe is not sufficient for movement. Crucially, though Valuation is dependent on an operation that operates strictly locally (Match), Valuation itself is not subject to locality. This means that non-local goals can value the probe under the right circumstances, and hence non-local goals can sometimes undergo movement.

By leveraging the decomposition of Agreement into two parts, we can get a handle on how probes navigate the competing pressures of locality and satisfying their needs. Match is subject to locality and Valuation is not. However, Valuation is subject to Best Match, which ensures that the goal will be valued as economically as possible. In the following pages, I show how this proposal can account for the Wh-over-Focus Preference of Mixtec.

2. A Problem of Locality

San Martín Peras Mixtec (SMPM, ISO: JMX) is an Otomanguean language spoken in the southern Mexican state of Oaxaca. All the data in this paper come from my own fieldwork with 6 speakers living in Mexico and California.

SMPM is a VSO language (Mendoza 2020), like other Mixtec languages (Macaulay 2005).

Both wh-phrases (11a) and foci (11b) undergo Ā-movement to a clause initial position. As I show in Hedding (2020), both types of constituents move to the same syntactic position, as has been argued for other languages (e.g. Rizzi 1997). In SMPM, this is evident based on the fact that these movements are in complementary distribution with one another, and the fronted wh-words and foci surface in the same position with respect to fronted quantifiers, temporal adverbs, and sentential negation.

1. This language is known by its speakers as Tu’un Ndá’vi or Tu’un Sávi
2. Mole is a catch-all term for several different sauces common in Oaxaca, which are typically made using a combination of chiles, nuts, and spices.
Given these sets of features and the relativization of the probe on C, derivations with both a wh-word and a focus provide an opportunity to investigate how Ā-movement probes navigate the potentially competing pressures of agreeing locally and completely satisfying the needs of the probe. In SMPM, when a wh-word and a focus both appear in the same clause, the wh-word must always front (13a), even if the focus is more local to the probe (13b). It is never possible to move a focus and leave a wh-word in-situ (13c).

(13) a. Yó shíní’í shìní’í_yí shìtí_a yá
   who brought — tortilla il_neut
   ‘Who brought the tortillas?’

b. Ná kíshashí Pedro rà
   what brought P. he
   ‘What did Pedro bring?’

c. *Pedro rà kíshashí ná
   P. he brought what
   Intended: What did Pedro bring?

The same basic preference to move wh-words over foci has been observed in several other unrelated languages, such as Hungarian (É. Kiss 1998: 16-17), Georgian (Borise & Polinsky 2018: 5) and for some speakers of Basque (Hualde & Ortiz de Urbina 2003: 495). This generalization is stated explicitly in (14).

(14) **The Wh-over-Focus Generalization:** In languages where wh-words and foci are in competition for a single position and only one can move, only wh-words will appear in that position when they co-occur within the same clause.

While the ungrammatical configuration in (13c) bears a superficial resemblance to a Focus Intervention Effect (e.g. Beck 2006), there are two properties of SMPM which suggest that the ungrammaticality of (13c) cannot be reduced to a constraint against a wh-word in situ that is c-commanded by a focus. First, contrastive focus can take DP-level scope in SMPM, as in English. This means that the wh-word need not be in the scope of the focus operator, the source of the Focus Intervention Effect according to Beck (2006). Second, there is an independent need for focus to move in the language, thus we don’t expect a moved focus in spec-CP to act as an intervenor for Agreement between C and a wh-word in situ, the source of Focus Intervention Effects according to Beck (1996) and Pesetsky (2000). Thus, the non-local movement of wh-words across foci presents an opportunity to understand the way that Ā-movement probes navigate competition between locality and probe satisfaction.

### 3. Proposal

I propose that the generalization in (14) can be attributed to a constraint which forces a probe to be valued by its “best match.” Specifically, I propose that SMPM has a probe on C that is relativized to the feature geometry [ufoc-uwh]. Merging this probe into the derivation initiates a local search for goals bearing these two features. In the following two subsections, I show how this probe and the constraint Best Match account for the movement of wh-words (§3.1) and foci (§3.2).

#### 3.1. Moving Wh-Words

Following work by Béjar (2003), I assume that a Match relationship can be established between a probe and a goal, as long as the goal bears some feature that entails the root of the probe. This means that
both wh-words and foci can match with the probe on C, given that they both bear the feature that is at the root of the probe: \[ \text{[FOC]} \].

I assume that Match is an operation that is subject to locality, thus in the case that a focus is structurally higher than a wh-word, the probe will match its \[ \text{[uFOC]} \] feature with the focus, as that is the most local \[ \text{[FOC]} \] in the derivation.

\[ \begin{align*}
\text{(15)} & \quad \text{Focus} > \text{WH} \\
& \quad \text{CP} \\
& \quad \text{c} \\
& \quad \text{[uFOC]} \\
& \quad \text{v} \\
& \quad \text{DP}_1 \quad \text{[FOC]} \quad 1 \\
& \quad \text{DP}_2 \quad \text{[FOC]} \quad \text{[WH]} \\
\end{align*} \]

After this first Match operation, there remains an unmatched \[ \text{[uWH]} \] feature on the probe. Following Bájar & Rezac (2009), I assume that a probe that has not found a match for one of its features after the first cycle of Agreement can probe again, potentially matching with a different goal. Thus, the unmatched \[ \text{[uWH]} \] feature will find its most local match: the structurally highest goal bearing a \[ \text{[WH]} \] feature.

\[ \begin{align*}
\text{(16)} & \quad \text{Focus} > \text{WH} \\
& \quad \text{CP} \\
& \quad \text{c} \\
& \quad \text{[uFOC]} \\
& \quad \text{v} \\
& \quad \text{DP}_1 \quad \text{[FOC]} \quad 1 \\
& \quad \text{DP}_2 \quad \text{[FOC]} \quad \text{[WH]} \\
\end{align*} \]

After these two cycles of Match, the probe has uniquely identified a set of goals which could potentially value its features.

\[ \text{(17) Set of Matched Goals (S): \{DP_{\text{FOC}}, DP_{\text{FOC-WH}}\}} \]

I propose that Valuation transfers features onto the probe, selecting from the set of goals that have matched with the probe. Furthermore, this transfer of features is subject to an economy constraint: Best Match.

\[ \text{(18) \quad \text{Best Match: A probe } P \text{ that has matched with a set of goals } S \text{ is valued by goal } G \text{ in } S \text{ such that the number of features on } G \text{ relevant to } P \text{ is greater than the number of features for any other } G' \text{ in } S.} \]

Crucially, I assume that a probe that triggers movement will move whichever goal(s) have valued it—simply matching with a probe does not trigger movement. That is, copying features onto a probe from a goal allows a copy of that goal to be Internally Merged as a specifier of the head that hosts the probe (cf. the composite operation move of Chomsky 2000). Taking the connection between copying features
onto the probe and Internally Merging a copy seriously, I assume that simply matching with a probe is not sufficient for movement.

(19) **Movement and Valuation:** A goal must value the probe in order to be attracted by that probe.

Due to the relativization of the probe in SMPM, a wh-word will always be a better match for the probe, as wh-words bear a superset of features borne by foci. This means that Best Match will always prefer moving a wh-word instead of a focus.

Under this view, Best Match is not simply a constraint which forces a probe to value as many of its features as possible. For probes that trigger movement, Best Match also forces a probe to economize its operations. Economizing over the copying of features—copying from the goal which will check off the most features all at once—simultaneously leads to economy of movement, since the only goals that are candidates for movement are those that have valued the probe. In the case of SMPM, a wh-word valuing a probe all at once is more economical than both the focus and wh-word partially valuing the probe, as this would in turn trigger multiple movement operations. Thus, Best Match can be related to other Economy Constraints which minimize operations triggered by probes, for instance, the following constraint proposed in Pesetsky & Torrego (2001).

(20) **Economy Condition**

A head H triggers the minimum number of operations necessary to satisfy the properties (including EPP) of its uninterpretable features.

Pesetsky & Torrego (2001): 359

Previous work has noted that more featurally specified goals can move across less featurally specified goals in some circumstances. Often, patterns of this type are accounted for representationally, such as with specific types of Relativized Minimality that are sensitive to feature geometries (Starke 2001, Rizzi 2004, Belletti & Rizzi 2013). Starke (2001), for instance, argues that a subclass can act as an intervener (21a), but a super class cannot (21b).

(21)  

(21)  

a. \(a...a\beta...a\)  

b. \(a\beta...a...a\beta\)  

Starke (2001): 8

This type of constraint—which is representational—is consistent with the derivational account presented here: a more featurally specified subclass (wh-words) can be attracted across a superclass (focus), but not vice versa.

3.2. **Moving Foci**

Recall that foci must move to the left periphery in SMPM when there is no wh-word present in the derivation (22a). It is ungrammatical for them to remain in situ (22b).

(22)  

(22)  

a. MARIA NÁ tśyā  __ shitā  

M. she made tortilla  

‘MARIA made the tortillas.’

b. *Tśyā MARIA NÁ shitā  

made M. she tortilla  

Intended: MARIA made the tortillas.

This is important for two reasons. First, it suggests that foci are not simply invisible to the probe on C. In the terms of Deal (2015), foci can interact with the probe. Second, it shows that foci can be attracted by the probe on C, just in case there are no wh-words in the derivation.

To see how the proposed analysis accounts for this pattern, consider a derivation that contains a focus but no wh-word. As before, the probe will match with the [foc] feature. Because the [uwh] of the probe remains unmatched, it can probe again. However, in this case, there is no goal bearing a [wh] feature,
so that feature remains unmatched. Following Preminger (2014), I assume that an unvalued feature must initiate a search, but that failing to find a match for an unvalued feature does not crash the derivation.

After the probe has matched all the features that it can, the goals that it has matched with form a set from which it will be valued. In this case, because the probe on C has only matched with one goal, the possible goals it could be valued by form a singleton set containing only the focused DP.

\[(\text{24}) \quad \text{Set of Matched Constituents (S): } \{\text{DP}_{\text{FOC}}\}\]

Thus, no modification needs to be made to the Best Match constraint proposed in (18)—because the probe has only matched with one goal, that goal is the only candidate to value it and move.

Importantly, this shows that in order to account for Wh-over-Focus Preference, Valuation of the probe must be delayed until all the features of the probe have attempted to find a match. If immediate Valuation were possible, then we would expect focused subjects to value the probe when they match with it. In the case where a focus in subject position is followed by a wh-word object, the [uwh] feature would then have to be valued by the wh-word. While the probe being valued by multiple goals is not problematic per se, we would then be forced to identify some criterion which would trigger movement of the wh-word over the focus. In other words, if we maintain the hypothesis that Valuation is a precondition for movement, then we would expect both the subject and the object—each of which has valued one of the two features of the probe—to be equally good candidates for movement. For this reason, I am committed to the possibility that Valuation can be delayed until after all features have attempted to find a match within their domain, at least in some languages (see §4.2).

The fact that Valuation is delayed until all Match relationships have been established is not readily apparent with morphological agreement. Given the nature of morphological agreement—as an independent Vocabulary Item—it is not apparent if the features necessary for any given Vocabulary Item are contributed by one, or more than one goal. Assuming that movement is a reliable diagnostic of which goal has valued the probe, investigating movement—in addition to morphological agreement—allows us to advance a more precise conception of Best Match, and of Valuation generally.

4. Consequences

4.1. Best Match is General

On the basis of Algonquin, Oxford (2014, 2019) argues that Best Match has a limited, specific role in the grammar. It is a mechanism that allows a probe to “decide” which of two equidistant goals it will Agree with. Under this conception, there is no tension between the need of a probe to match all of its

---

\[3\] Béjar & Rezac (2009) (pg. 50-51) discuss the possibility that one valuation might overwrite another in cases where the first valuation is a proper subset of the second. Extending this logic to the phenomenon under discussion, we could say that the focus values [uFOC], then the wh-word values it again along with valuing [uWH], overwriting and mooting the first valuation. While this analysis is possible, it would require an additional stipulation that valuation does not immediately trigger Internal Merge.
features and Agree with the most local goal in its c-command domain—probes Agree locally, but if there are two goals that are equally local, then Best Match makes the final call.

While this conception is consistent with the facts of Algonquin, it cannot be easily extended to Mixtec. Unlike in Algonquin, in Mixtec the subject asymmetrically c-commands the object, as evidenced, for instance, by the fact that condition-A anaphors are licit in object position but not in subject position. However, despite this difference, the basic pattern of Agreement that the two languages display closely mirror one another: a probe will Agree with the goal that most completely values its features.

Thus, expanding our scope beyond Algonquin—and to patterns of Ā-movement—reveals that Best Match and equidistance are orthogonal to one another. In languages where multiple goals are equidistant from some probe, Best Match can determine which will ultimately value the probe. This is the case in Algonquian. However, as the data from Mixtec demonstrates, equidistance is not a prerequisite for Best Match to play a role in deciding which of two goals that have matched with a probe will ultimately value it. Consequently, I propose that Best Match is a general constraint that applies anytime that a probe matches with more than one goal. Crucially, a probe cannot simply “look ahead” to see if a better goal is coming. Instead, all and only the goals which have entered into a Match relationship with the probe are candidates to value the probe.

This conception of Best Match builds on previous work by Coon & Bale (2014) and van Urk (2015) which argue that Best Match can trigger Agreement with a non-local goal. However, unlike previous proposals, the conception of Best Match that I advance here offers a derivational account for this preference using familiar operations—Match and Value—without giving probes the ability to look ahead. In this way, it offers a more precise conception of how a probe navigates the competing pressures of locality and satisfaction.

4.2. Early vs. Late Valuation

A key component of the proposal advanced here is that Valuation of a probe can be delayed until all of its features have attempted to Match with some goal. This delay in Valuation is apparent in cases where some goal that can value the probe (such as a focus), doesn’t, just in case there is a non-local goal that constitutes a better match. This differs from recent work by Coon et al. (to appear), who argue that a probe will immediately be valued by any goal that it matches with. Crucially, for Coon et al. Valuation occurs even if the goal can only value a subset of the features that the probe is relativized to.

This difference in derivational timing is integral to their analysis of the restriction on Ā-extraction of ergative arguments in Mayan languages. They propose that the probe on C responsible for Ā-extraction is relativized to search for [ud] and [uĀ] features. Furthermore, they assume that the absolutive argument moves to a position that is structural higher than the ergative argument. In derivations with an ergative argument that could be Ā-extracted, it will be structurally lower than an absolutive argument. When the absolutive argument is a DP, the probe on C initiates a search and matches its [d] feature with the absolutive argument. Crucial to their analysis is the proposal that this match relationship immediately triggers a valuation of the [d] feature. Then, when the [ā] on the probe initiates a search, it matches with the ergative subject, and is also valued by it. Thus, in derivations where the ergative argument could theoretically be Ā-extracted, the probe on C has matched with and been valued by two different goals.

(25) “Feature Gluttony” in Ā-Probing

\[
\text{Coon et al. (to appear): 18}
\]

According to Coon et al. (to appear), it is this configuration of multiple valuations that make Ā-extraction of ergative arguments ungrammatical. Specifically, building on a proposal from Coon & Keine (2020), they argue that every goal that values the probe must move.

(26) If a segment of a movement-inducing probe on a head H has agreed with an XP, this XP must
undergo movement to the specifier of H.

Coon et al. (to appear): 20

Thus, if valuation occurs immediately, both the subject and object will value the probe, leading to conflicting demands. Namely, both the subject and object will need to move, but, according to Coon et al. (to appear), they can’t move sequentially or simultaneously. These conflicting demands cause the derivation to crash, preventing Ā-extraction or ergative arguments.

A crucial element of their analysis, then, is that a probe will immediately be valued by a goal that it matches with, even if it is only a partial match. This is distinct from the the analysis proposed here, which relies on the possibility for late valuation in order to account for the Wh-over-Focus pattern of Mixtec. Straightforwardly adopting the analysis of Coon et al. (to appear) would incorrectly predict that a wh-word could not move across a focus, contrary to fact. However, straightforwardly applying the Best Match constraint proposed here to Mayan would incorrectly predict that ergative extraction is possible in those languages, given the relativization of the probe that Coon et al. (to appear) propose. While I leave for future work a more detailed exploration of the timing of Valuation cross-linguistically, I suggest that these conflicting accounts demonstrate that Ā-extraction may be a fruitful domain to further investigate the precise way that articulated probes are valued.

5. Conclusion

Probes have multiple needs, and sometimes these needs come into conflict with one another. In this paper, I have identified a pattern of Ā-movement in Mixtec which instantiates a tension between local Agreement, and completely satisfying the needs of the probe.

The proposal advanced here builds on previous work, primarily in the domain of morphological agreement, which suggests that a probe will Agree with the goal that constitutes its best match. In order to account for the pattern in Mixtec, I propose an articulated probe on C which enters into local matching relationships in cycles. Once matching is complete, the probe is valued by the goal it has matched with that satisfies the most features it is relativized to. Given the relativization of the probe and the respective features of foci and wh-words, wh-words will always constitute a better match for the probe, and thus will be preferentially moved.

In essence, this account analogizes the Wh-over-Focus Generalization to a person hierarchy effect—in Algonquin, morphological agreement with (more featurally specified) local persons is preferred, in Mixtec movement of (more featurally specified) wh-words is preferred. If this analysis is on the right track, then we might expect to find analogies of familiar person hierarchy effects might be found within the Ā-domain of the world’s languages. For instance, beyond the Wh-over-Focus Effect, we might find languages which restrict the order of wh-words and foci in the left-periphery when they both extract. Additionally, we might expect to find languages that use morphology to signal the order of wh-words and foci within the clause, analogous to direct and inverse morphology (e.g. Aissen 1997). In sum, extending what we know about Agreement in the A-domain to the Ā-domain—and expanding the scope of our search—gives us the opportunity to advance the investigation of the precise workings of Agreement.

References


4 See Erlewine (2018) pgs. 668-669 for a possible example of this type of ordering effect in Toba Batak.